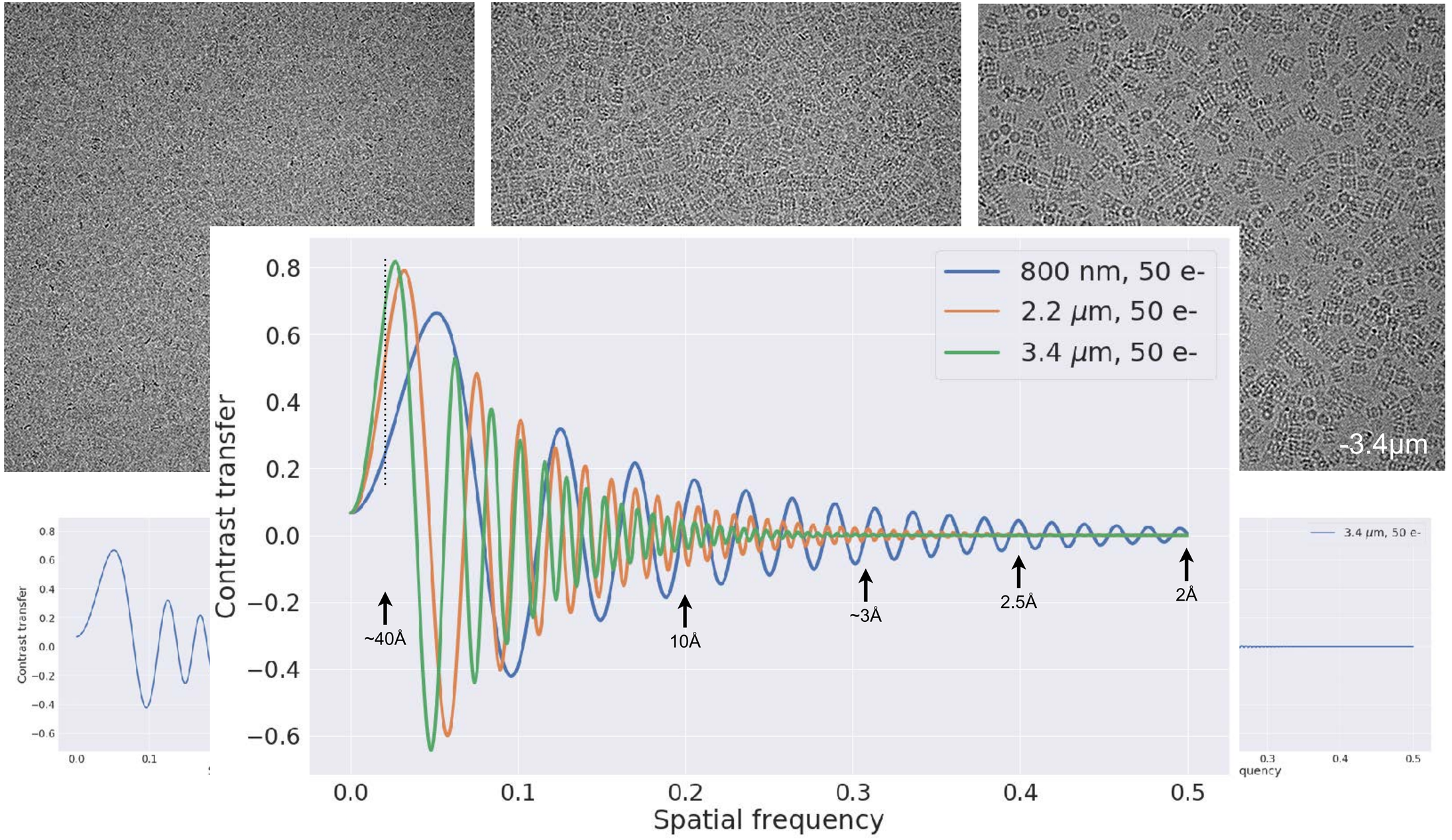
Influence of CTF on image



Spherical aberration: C_s

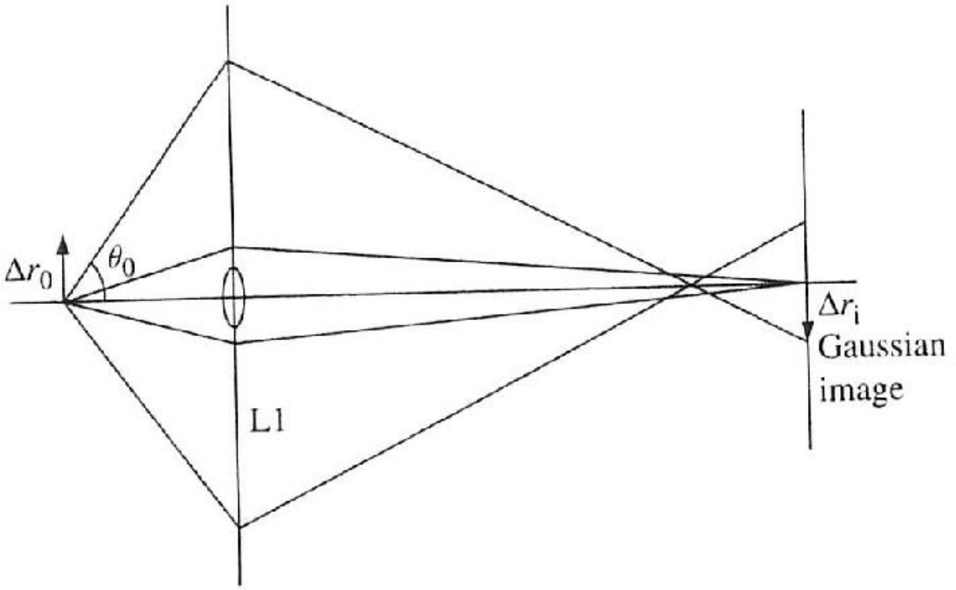


Figure 2.13 The effect of spherical aberration. Rays passing through the outer zones of the lens (far from the axis) are refracted more strongly than those paraxial rays passing close to the axis. These outer rays meet the axis before the Gaussian image plane, and meet that plane a distance Δr_i from the axis. Spherical aberration (or aperture defect) is the most important defect affecting the quality of high-resolution images.

From John Spence: High-resolution electron microscopy

Chromatic aberration: C_c

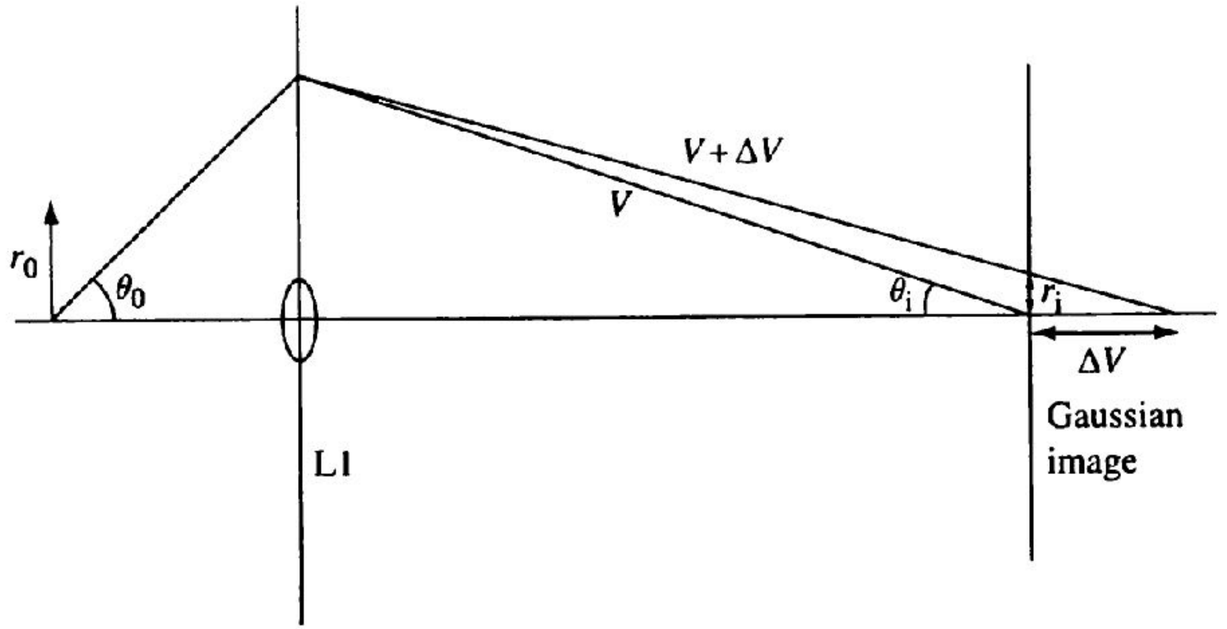


Figure 2.14 The effect of chromatic aberration. The faster electrons which have been accelerated through a potential $V + \Delta V$ are less strongly refracted than lower-energy electrons accelerated through a potential V. These higher-energy electrons are thus brought to a focus beyond the Gaussian image plane, which they pass at a distance r_i from the axis.

From John Spence: High-resolution electron microscopy

Astigmatism

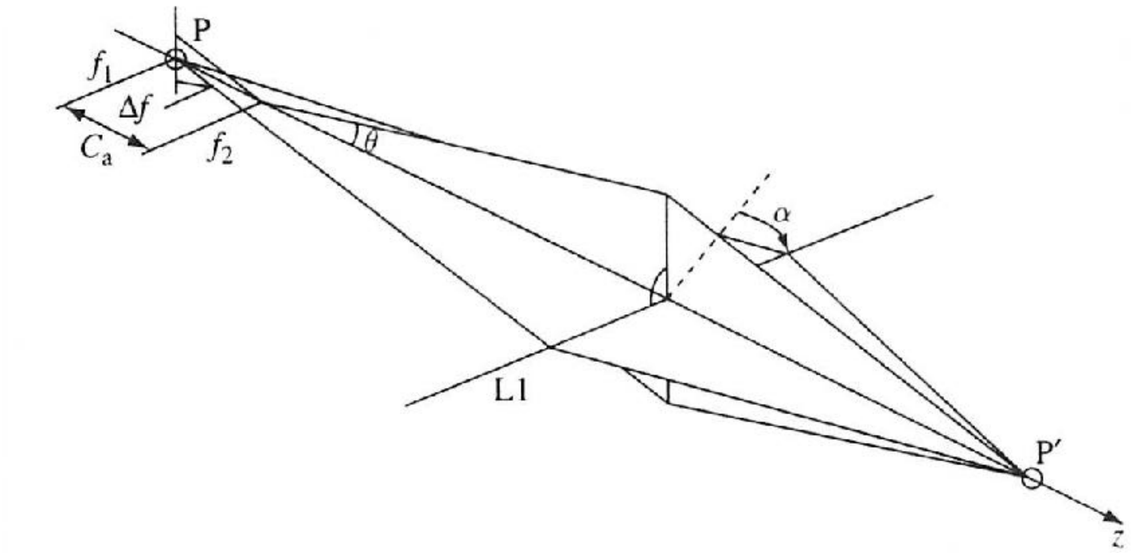
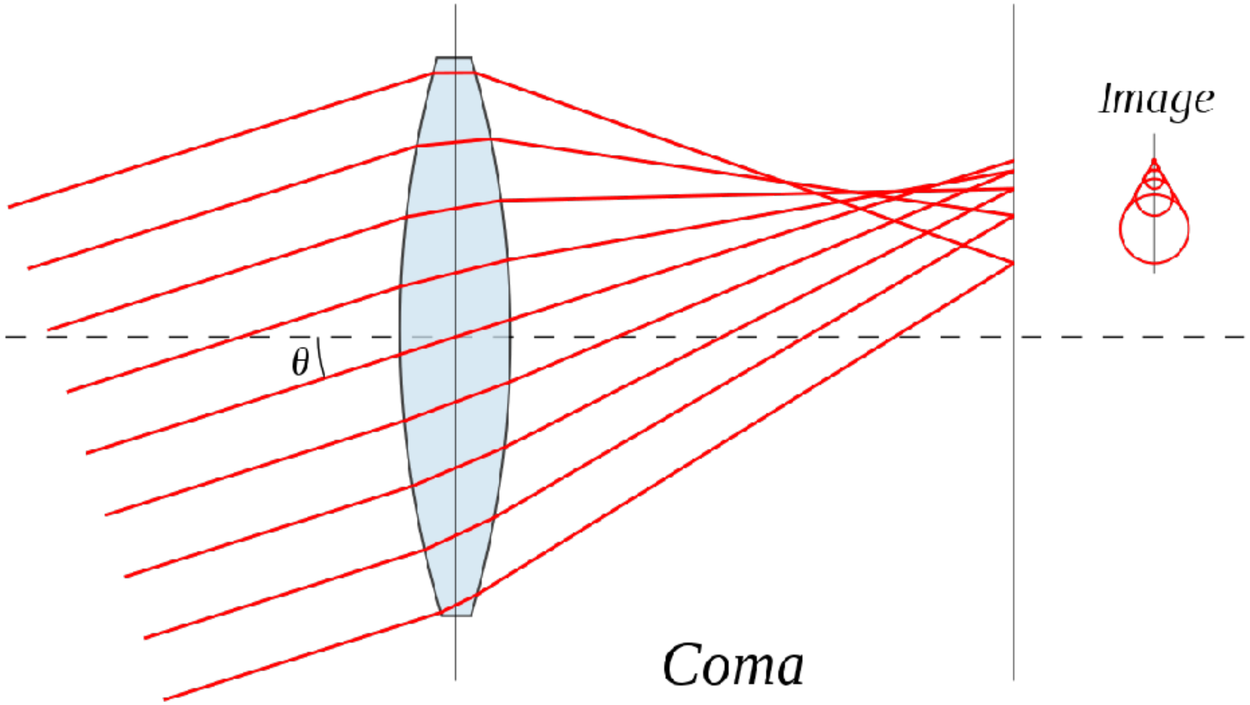
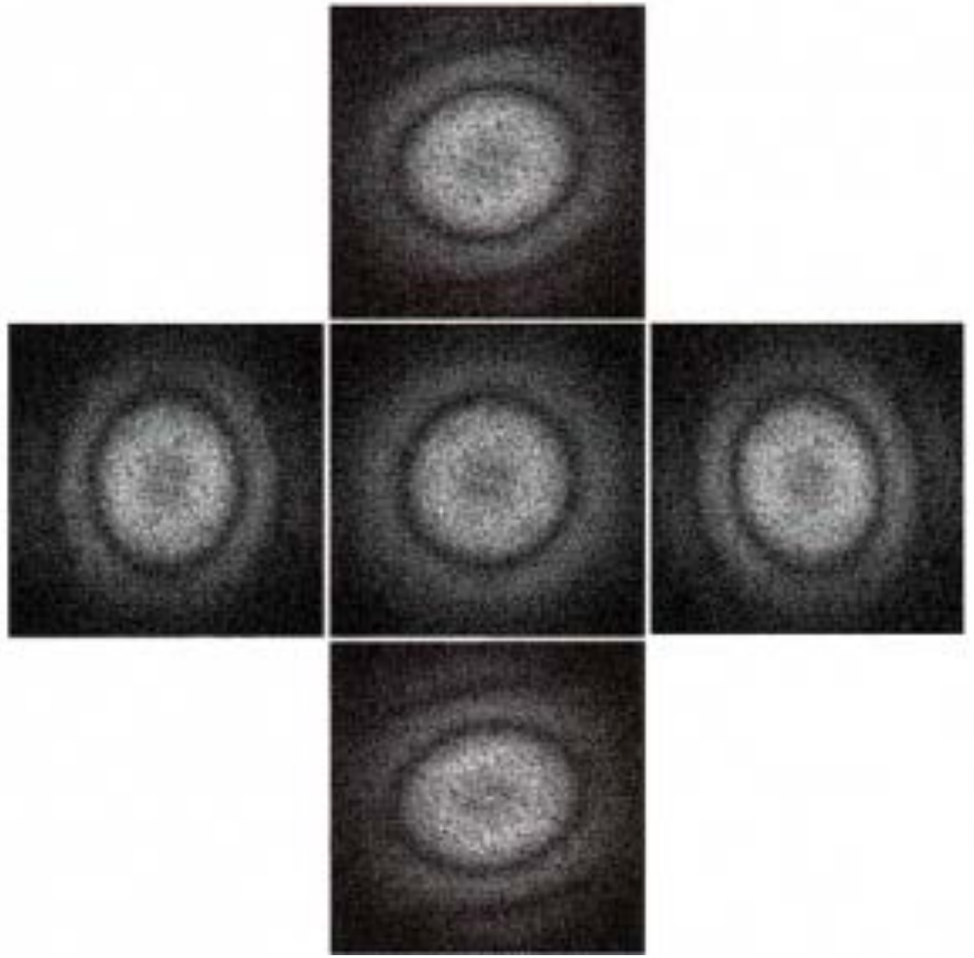


Figure 2.15 Astigmatism. The focal length of the lens depends on the azimuthal angle α of a ray leaving the object. These rays are in a plane containing the optic axis. Planes at right angles for the maximum and minimum focal length are shown, with a mean focus f. The difference between the maximum and minimum focus is the astigmatism constant C_a.

From John Spence: High-resolution electron microscopy

Coma





“You just *look at* the thing!”

Richard Feynman: There’s plenty of room at the bottom
(December 29, 1959, lecture to American Physical Society):

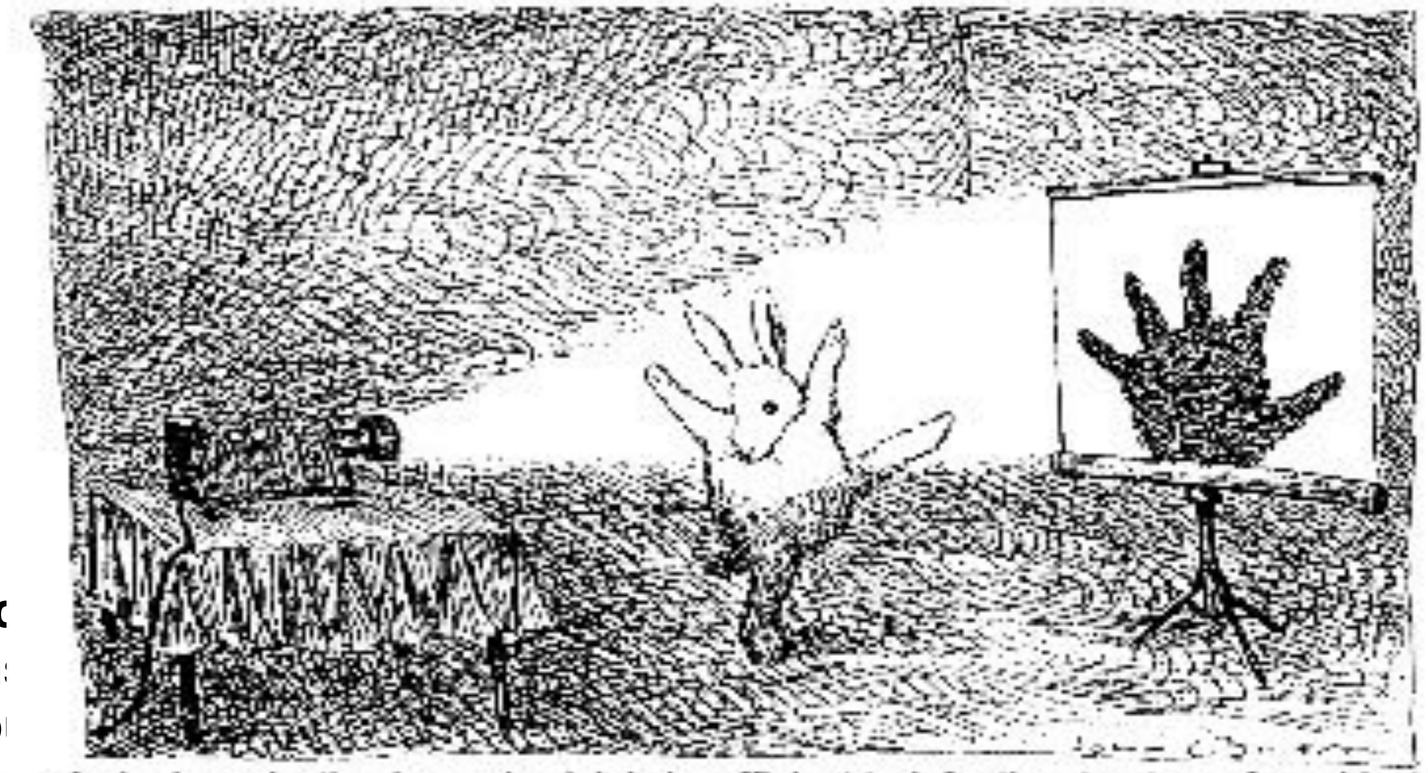
“It is very easy to answer many of these fundamental biological questions: you just *look at the thing*!”

“Unfortunately, the present microscope sees at a scale which is just a bit too crude. Make the microscope one hundred times more powerful, and many problems of biology would be made very much easier.”

“… the biologists would surely be very thankful to you”

Reconstructing 3D object from 2D projection images

Central Sec
Fourier tran
section thro
the directior



A single projection image is plainly insufficient to infer the structure of an object.
John O'Brien; © 1991 The New Yorker Magazine

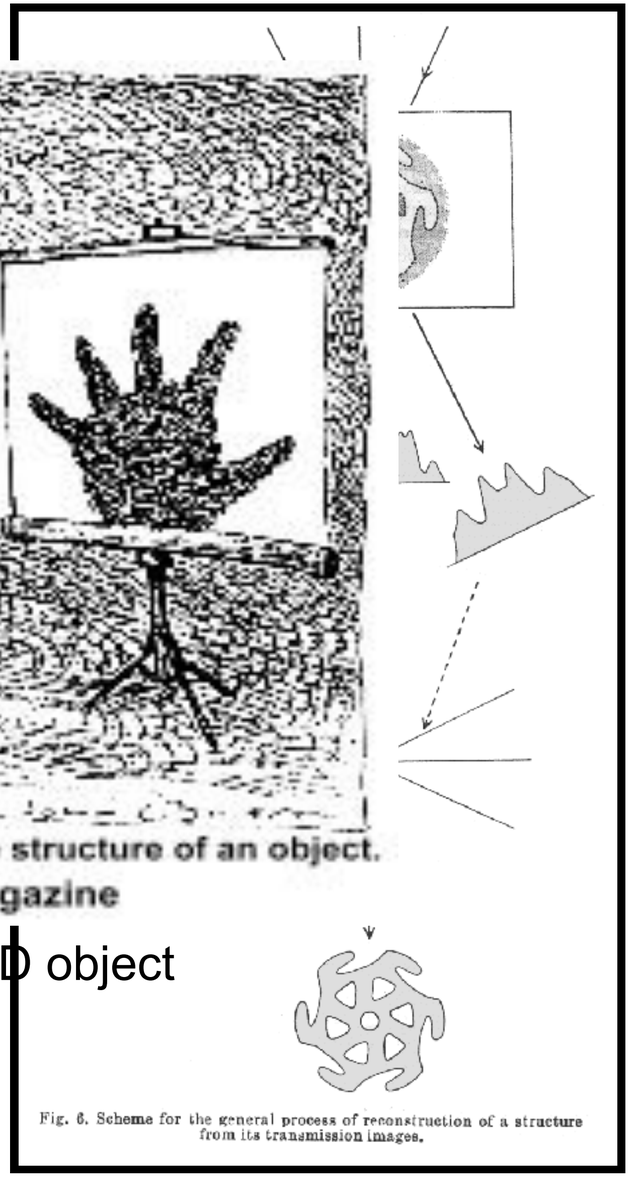


Fig. 6. Scheme for the general process of reconstruction of a structure from its transmission images.

DeRosier and Klug (1968)

DeRosier, D. and Klug, A. (1968) "Reconstruction of three dimensional structures from electron micrographs" *Nature* ***217*** 130-134

Images are 2D projections of a 3D object

Molecular electron microscopy of biological sample

Strong electron scattering power means two things:

1 high vacuum of microscope column;

2 strong scattering with protein sample;

Problems:

1 dehydration of biological sample;

2 radiation damage by high energy beam;

Molecular electron microscopy of biological sample

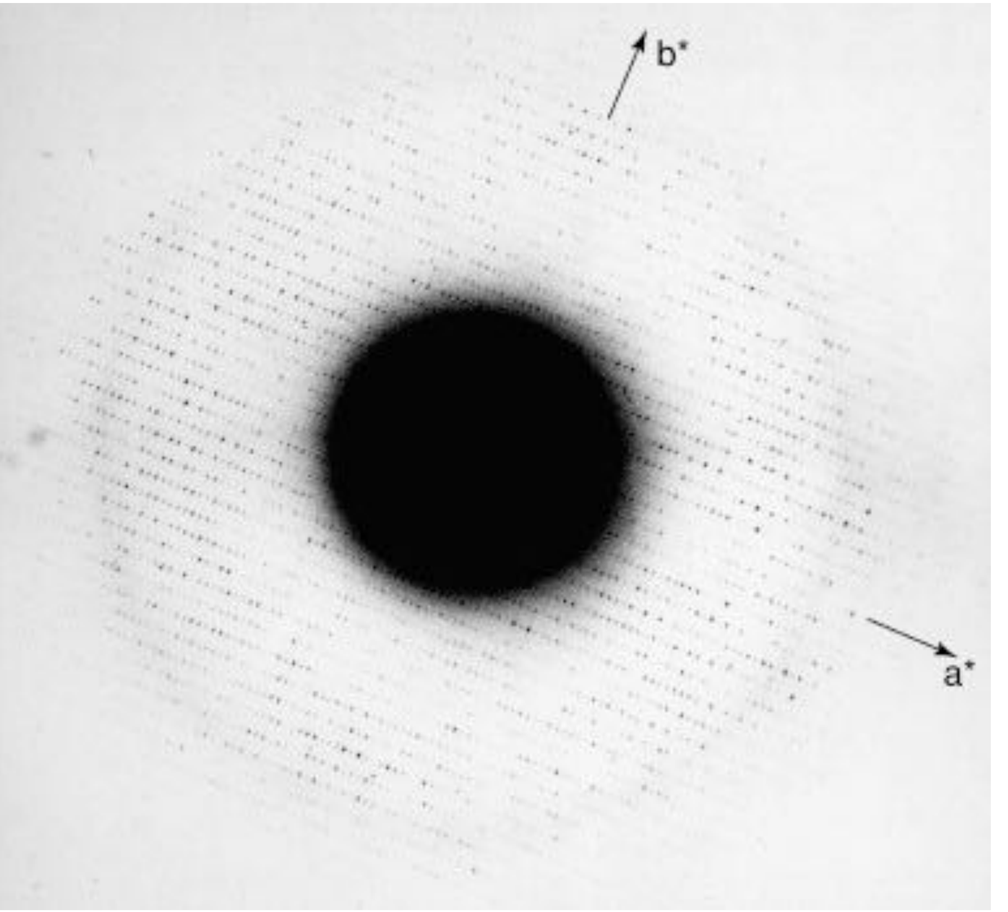
Strong scattering by high-energy electrons imposes two challenges to biological samples:
- dehydration caused by high vacuum within electron microscope column destroys biological samples;
- severe radiation damage caused by high-energy electron beam destroys biological samples;

* Shadow casting (Williams & Wycoff, 1945);
* Positive staining (Pease & Baker, 1948);
* Glass knives for microtomy (Hartmann & Latta, 1950);
* Diamond knives (Fernandez-Moran, 1953);
* Negative staining (Hall, 1955);

Frozen hydration preserve structural integrity to atomic level.

Taylor K and Glaeser RM (1974) “Electron diffraction of frozen, hydrated protein crystals” *Science* **186**, 1036-1037

Taylor and Glaeser (2008) “Retrospective on the early development of cryoelectron microscopy of macromolecules and a prospective on opportunities for the future” Journal of Structural Biology



Cryo-electron microscopy

Against dehydration:

glucose/trehalose embedding: using glucose to substitute water, thus maintain hydration in the high vacuum. Only used for 2D crystal;

Frozen hydration: using plunge freezing to avoid crystal ice. Mostly for single particle;

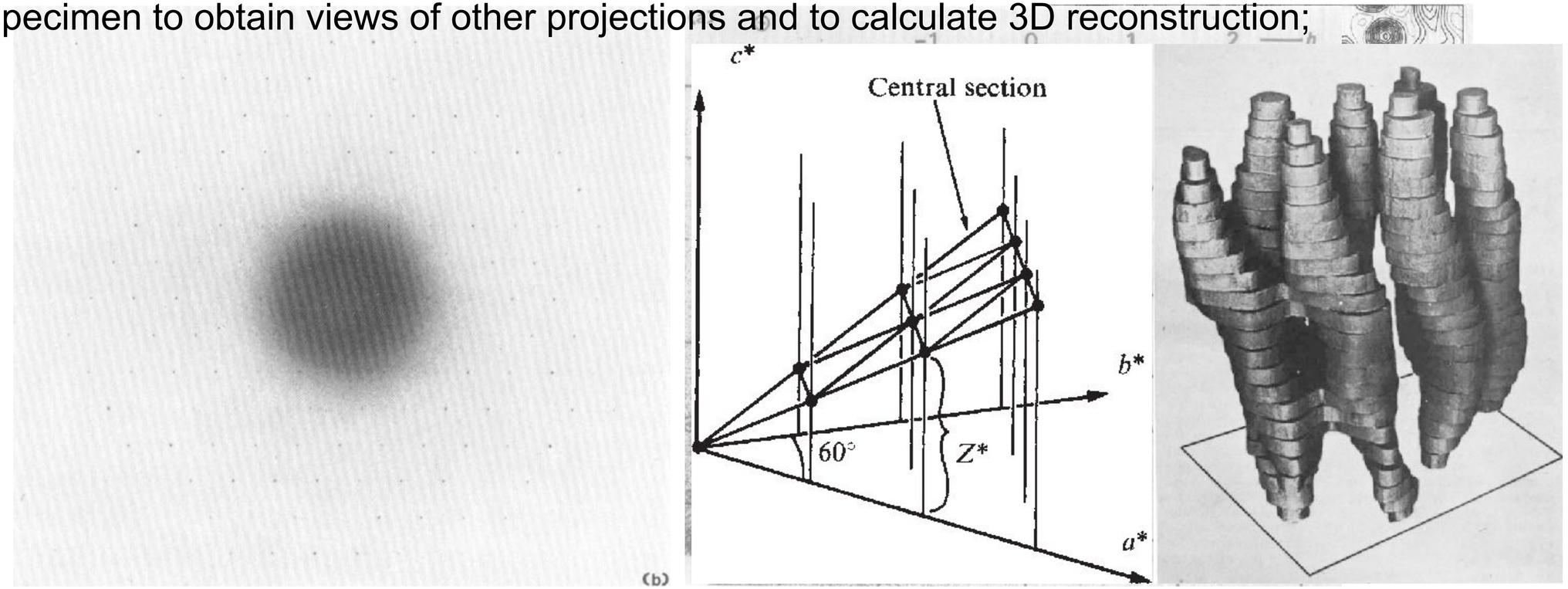
- Against radiation damage:

  Low-temperature: LN2 (~80K) or LHe (~10K); Challenges to the instrumentations;

  Low-electron dose: Low-dose imaging; Results in extremely noisy images, challenges for the data processing;

Structure of unstained crystalline specimen by electron microscopy

- Substituting water with sugar to prevent dehydration;
- Using crystalline samples to obtain sufficient signals from images recorded with low electron dose;
- Tilting specimen to obtain views of other projections and to calculate 3D reconstruction;



Electron diffraction (left) and Fourier transform from image of Bacteriorhodopsin

Henderson R and Unwin N (1975) "Three-dimensional model of purple membrane obtained by electron microscopy" *Nature* **257**, 28-32.

Unwin N and Henderson R (1975) "Molecular structure determination by electron microscopy of unstained crystalline specimens" *Journal of Molecular Biology* **94**, 425-440.

Single particle EM: averaging of low dose image of non-periodic objects

J Frank (1975) “Averaging of low exposure electron micrographs of non-periodic objects” *ultramicroscopy* ***1,*** 159.

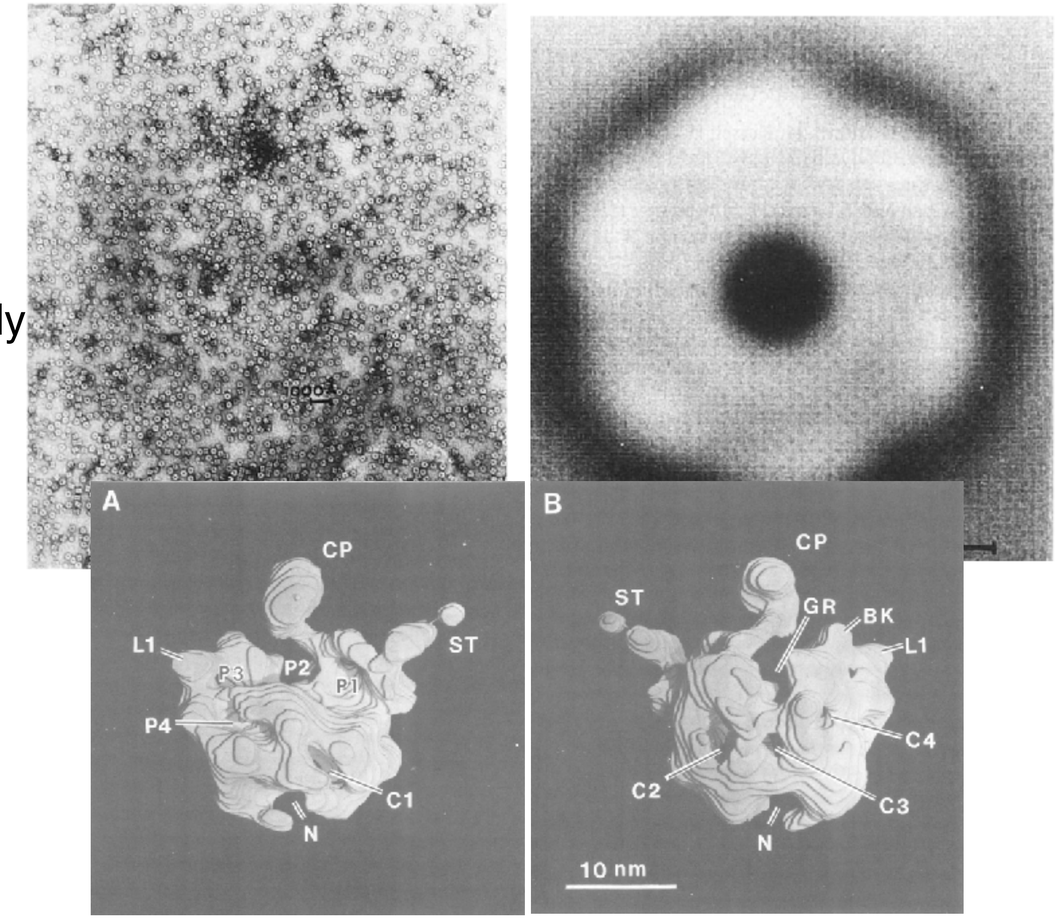
“We will investigate how the average techniques could be extended to this general case. Of all the possible regular specimen …., we are interested in those which form identical particles, sufficiently well separated on the microscope grid so as not to overlap.”.

Frank, J. Goldfarb, W, Eisenberg, D. and Baker, T.S. (1978) “Reconstruction of glutamine synthetase using computer averaging” *ultramicroscopy* ***3,*** 283-290.

“A single low-dose micrograph of a maximally tilted specimen will supply all the Fourier information contained in a cone up to that tilt angle”.

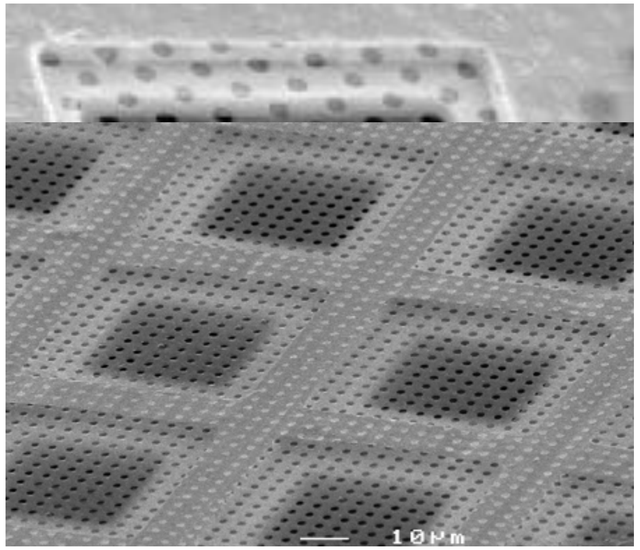
Radermacher, M., Wagenknecht, T., Verschoor, A., and Frank, J. (1987) “Three-dimensional Structure of large ribosomal subunit from *Escherichia coli*” The *EMBO Journal* ***6,*** 1107-1114.

E. coli 50S ribosome by random conical tilt (RCT)

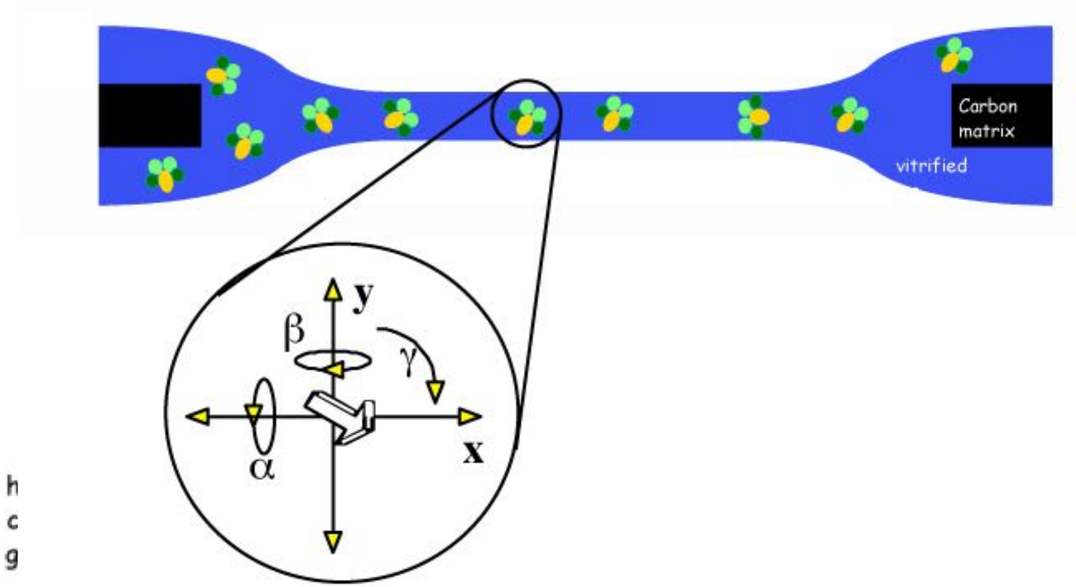


Frozen hydrated specimen preparation for single particle cryo-EM

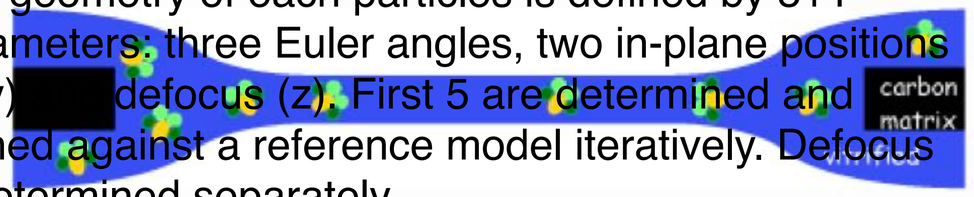
Adrian M, Dubochet J, Lepault J & McDowall AW (1984)
Cryo-electron microscopy of viruses. *Nature* **308**, 32-36.

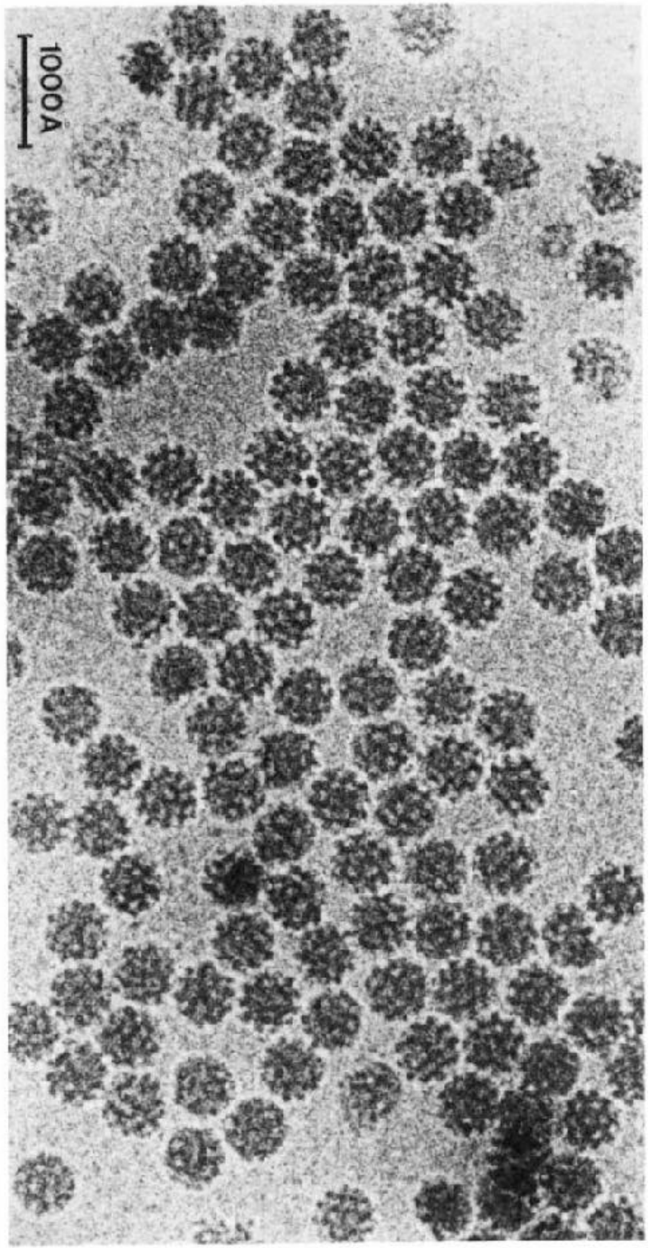


Quantifoil grid

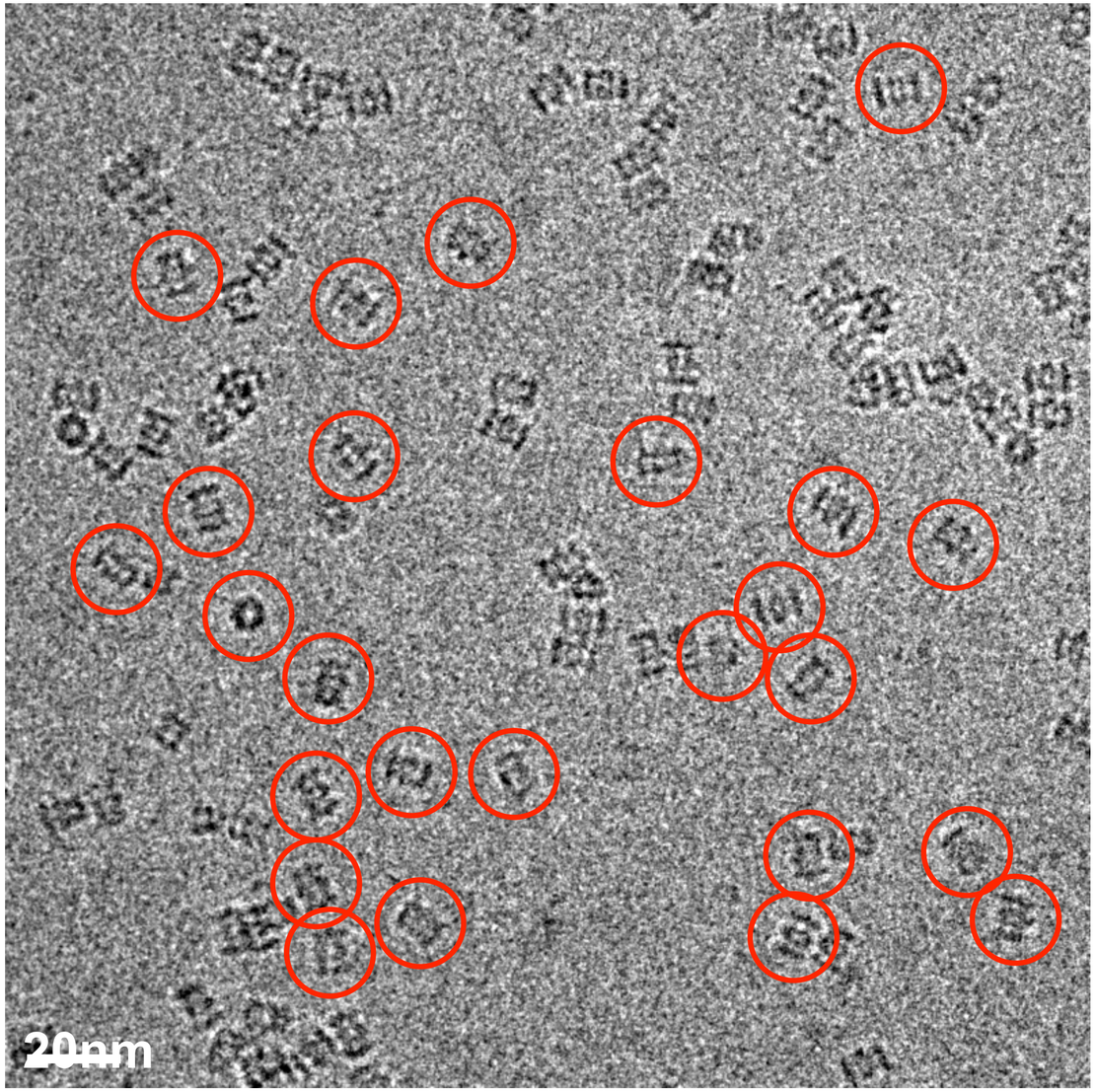


The geometry of each particles is defined by 5+1 parameters: three Euler angles, two in-plane positions (x, y) and defocus (z). First 5 are determined and refined against a reference model iteratively. Defocus is determined separately.





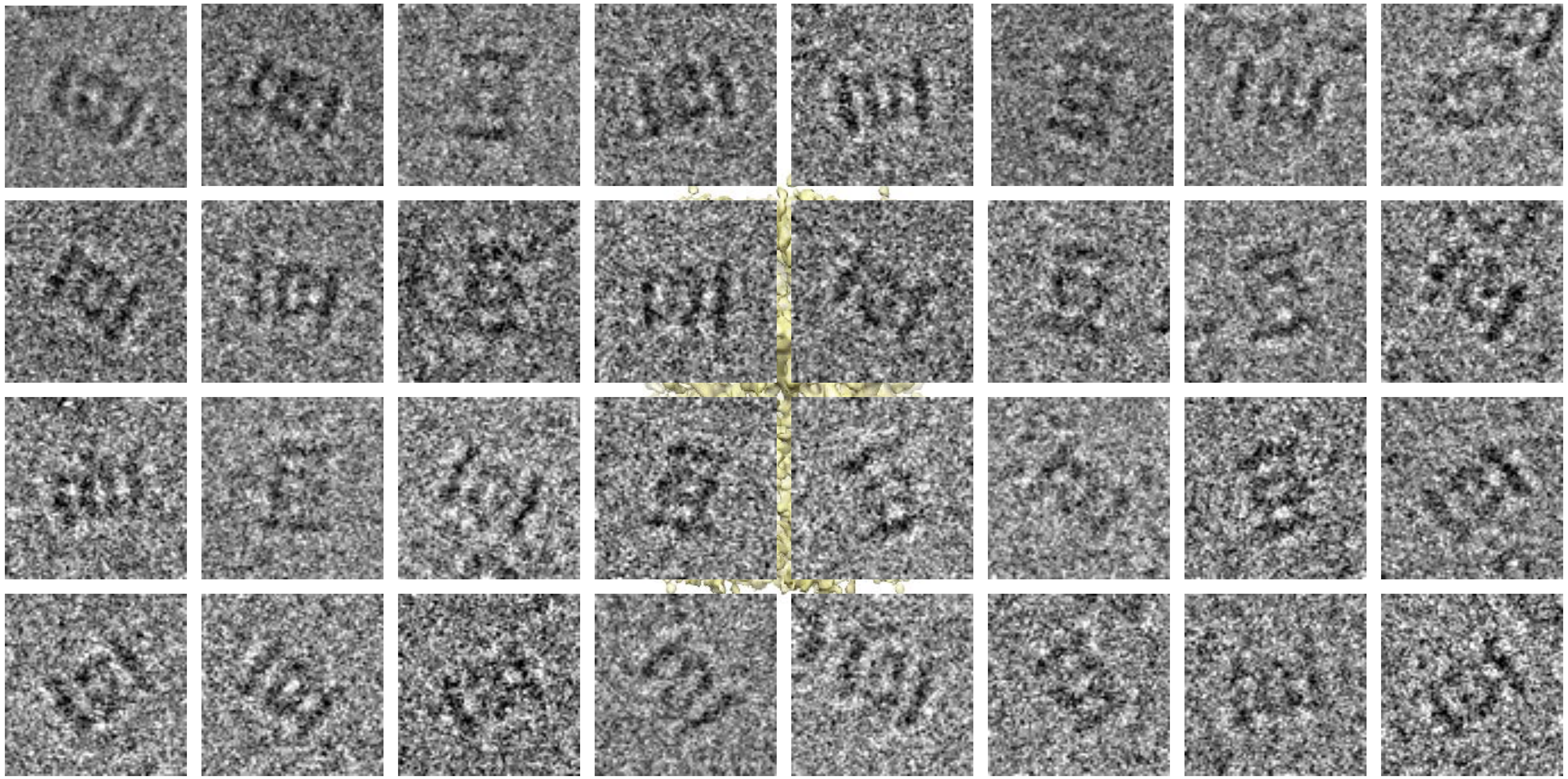
Single particle cryo-EM



Cryo-EM image of frozen hydrated archaeal 20S proteasome

Selecting particle images

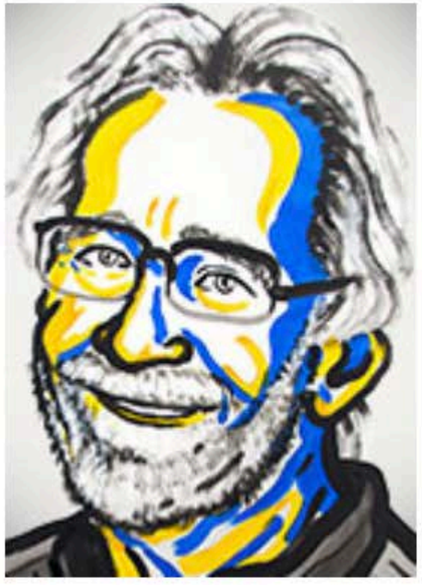
Single particle cryo-EM



3D reconsBoxtioun (hancsividuaalignmenticle imagestion, 2005)

Rabl, Smith, Yu, Chang, Goldberg and Cheng (2008) ***Molecular Cell***

The Nobel Prize in Chemistry 2017



© Nobel Media. Ill. N. Elmehed

Jacques Dubochet

Prize share: 1/3

© Nobel Media. Ill. N. Elmehed

Joachim Frank

Prize share: 1/3

© Nobel Media. Ill. N. Elmehed

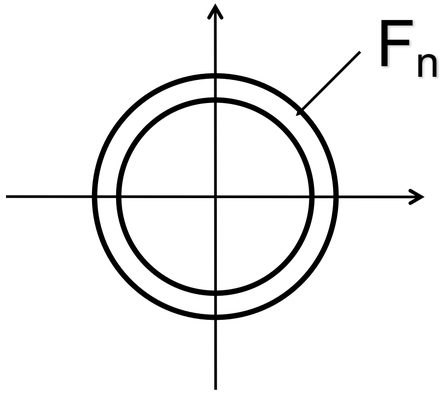
Richard Henderson

Prize share: 1/3

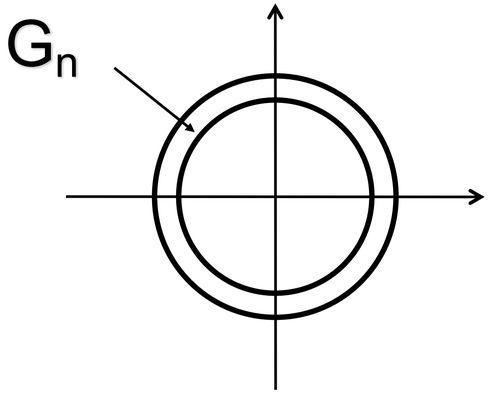
The Nobel Prize in Chemistry 2017 was awarded to Jacques Dubochet, Joachim Frank and Richard Henderson *"for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution"*.

Resolution estimation

In single particle cryoEM the resolution is often estimated by Fourier Shell Correlation.



Reconstruction 1



Reconstruction 2

$$FSC(R) = \frac{\sum_{n \in R} F_n G_n}{\left\{ \sum_{n \in R} |F_n|^2 \sum_{n \in R} |G_n|^2 \right\}^{1/2}}$$

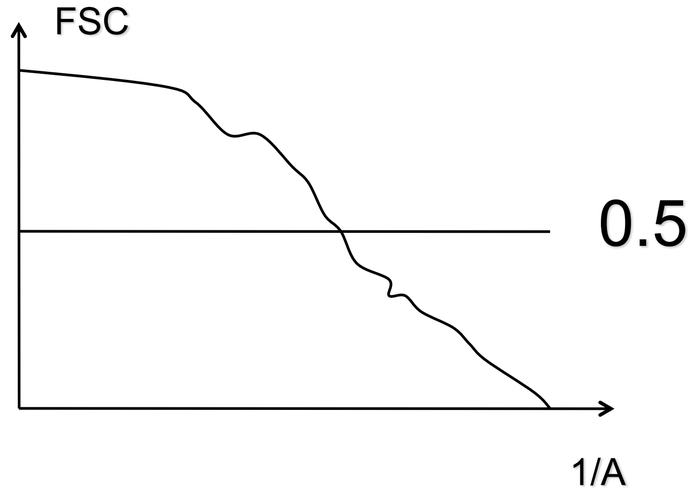
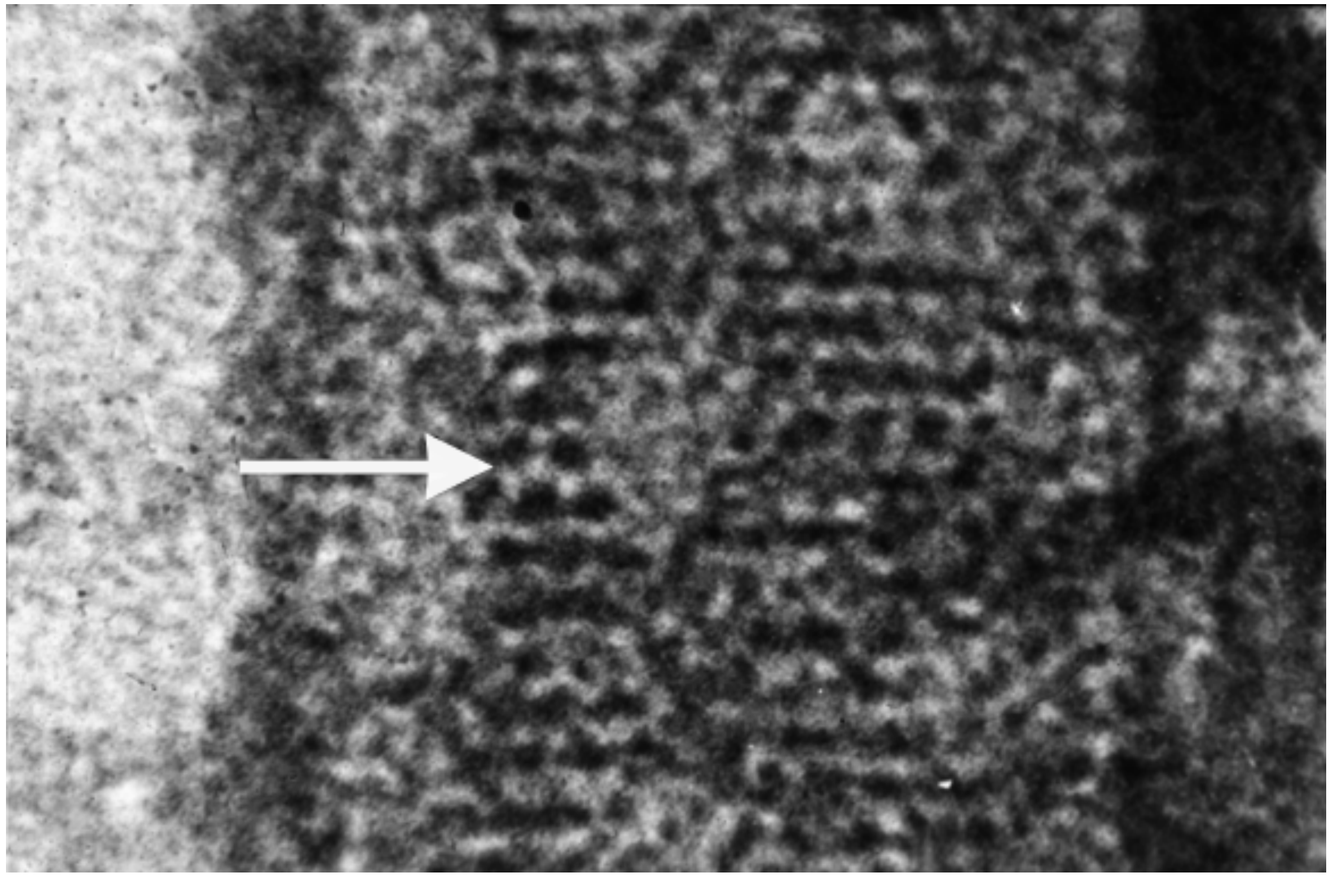
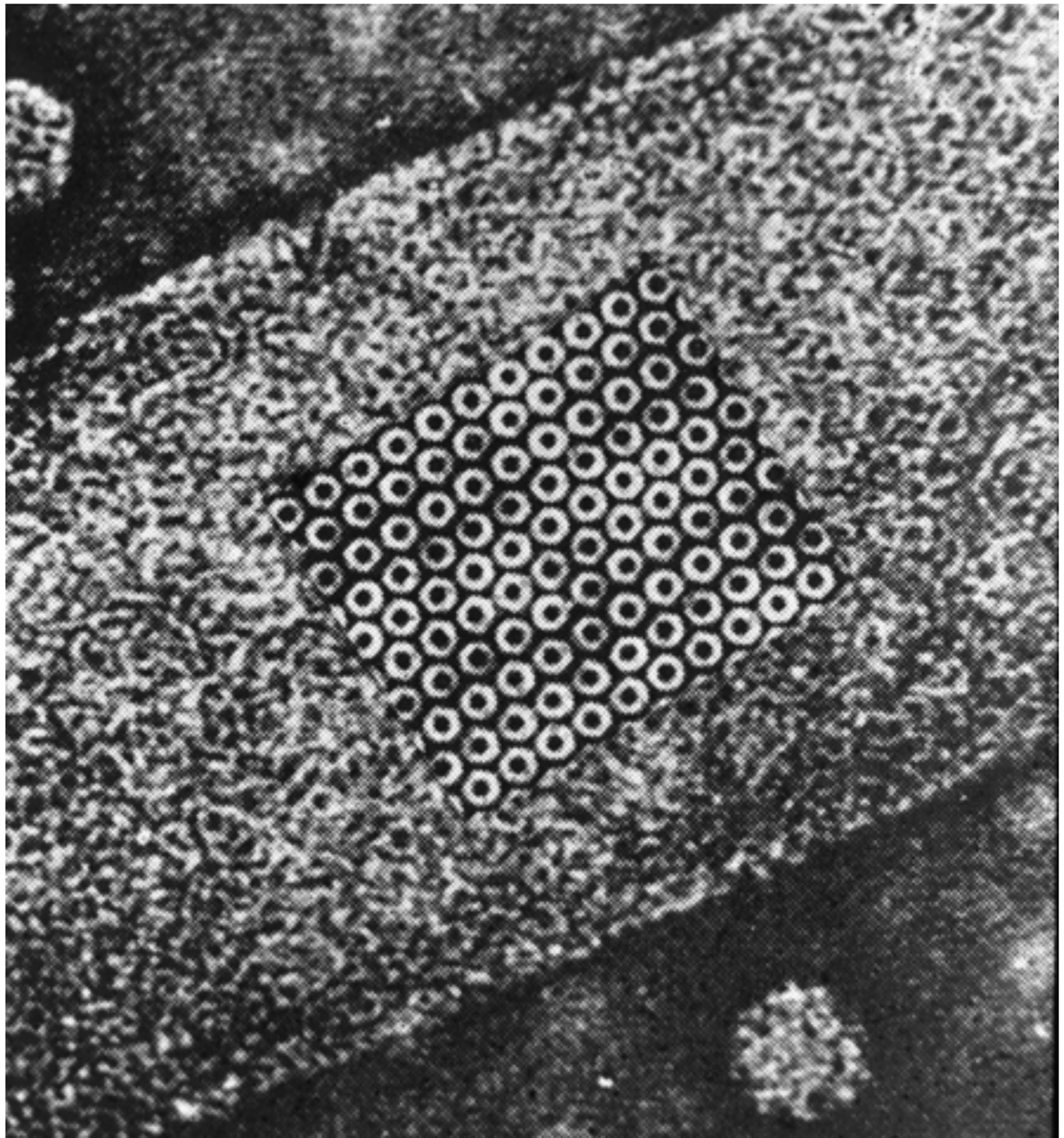


Image averaging

Cryo-EM images are very noisy; have extremely low signal-to-noise ratio. Averaging of a large number of images are necessary to improve the SNR.



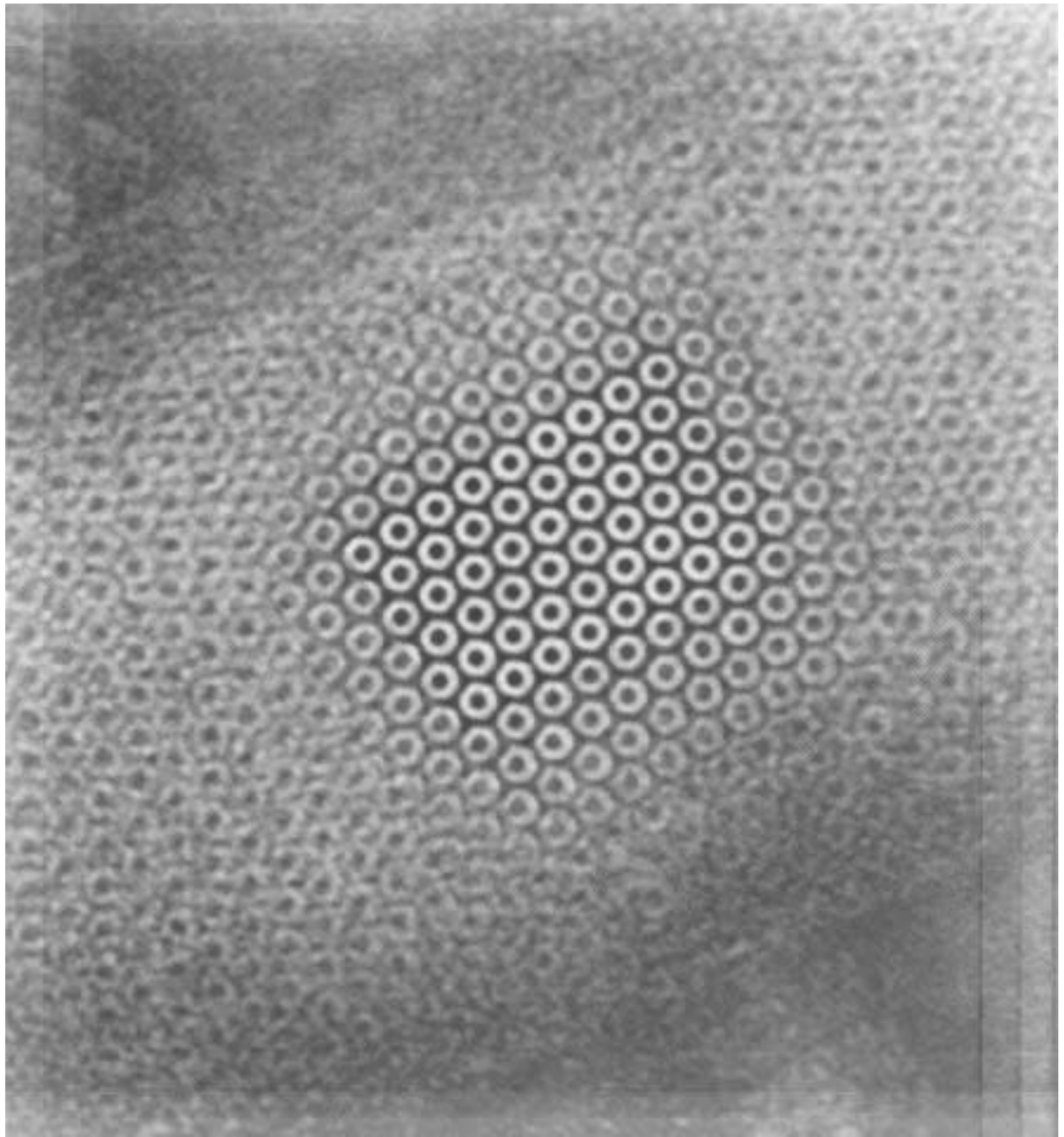
Averaging in darkroom



Photographic image superposition (averaging) by Roy Markham, who shifted image and added to the original in darkroom.

The trick is to know decide much and which direction to shift the image for superposition.

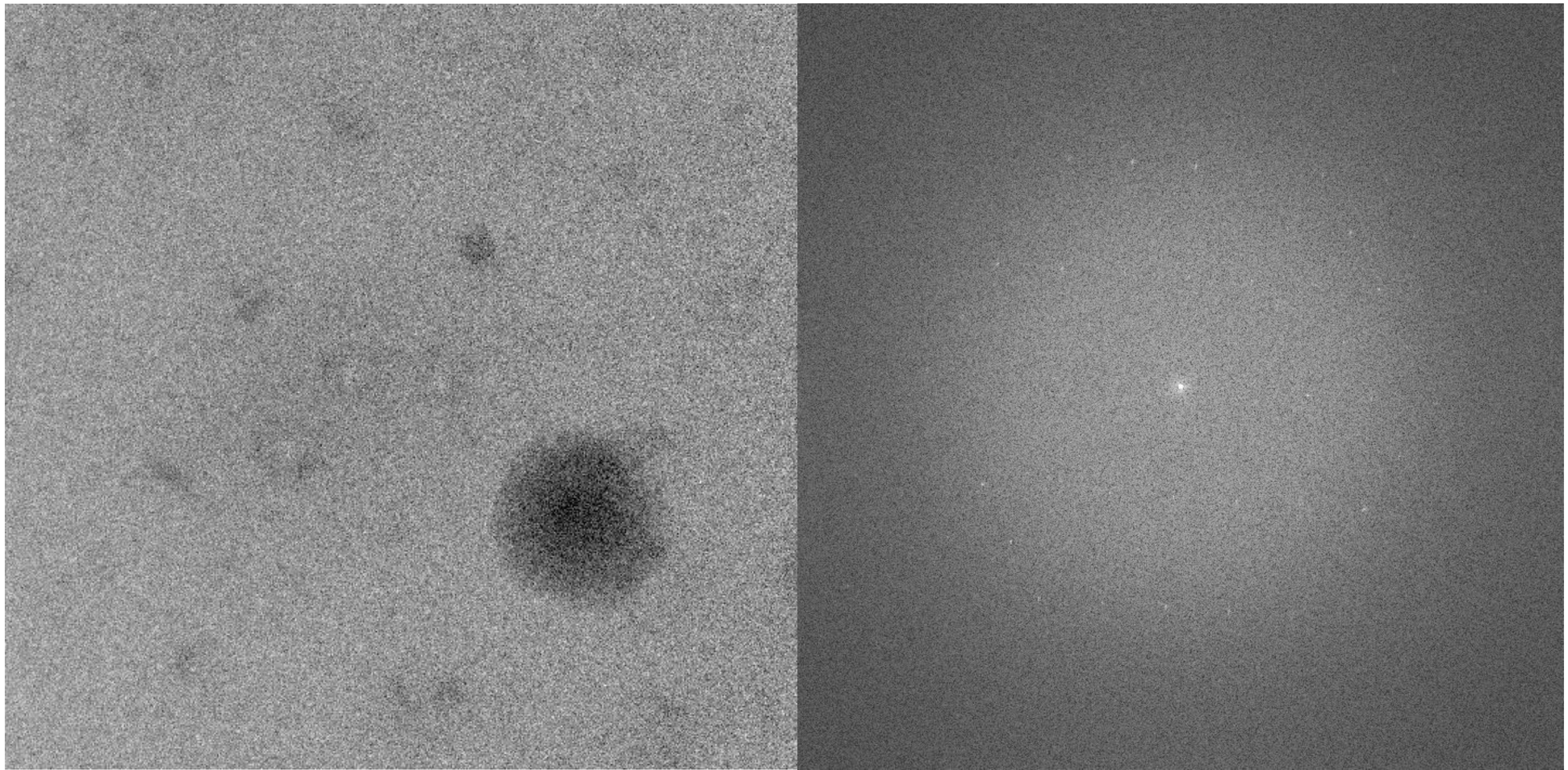
Averaging in computer.

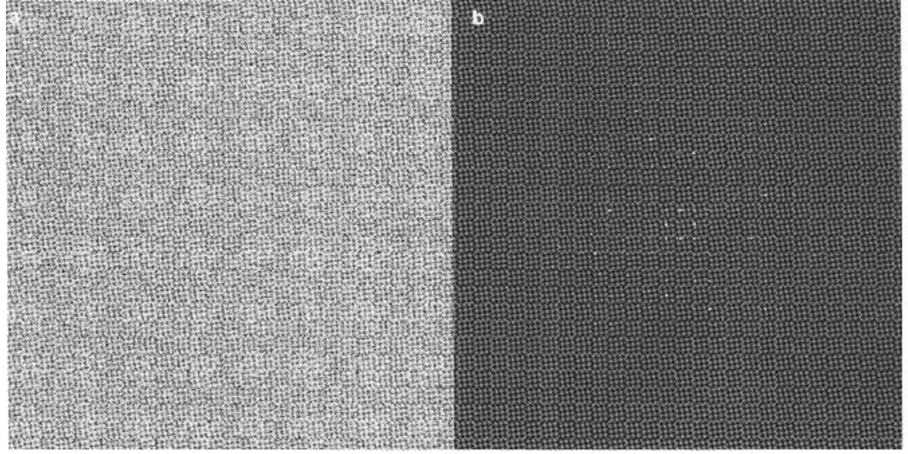


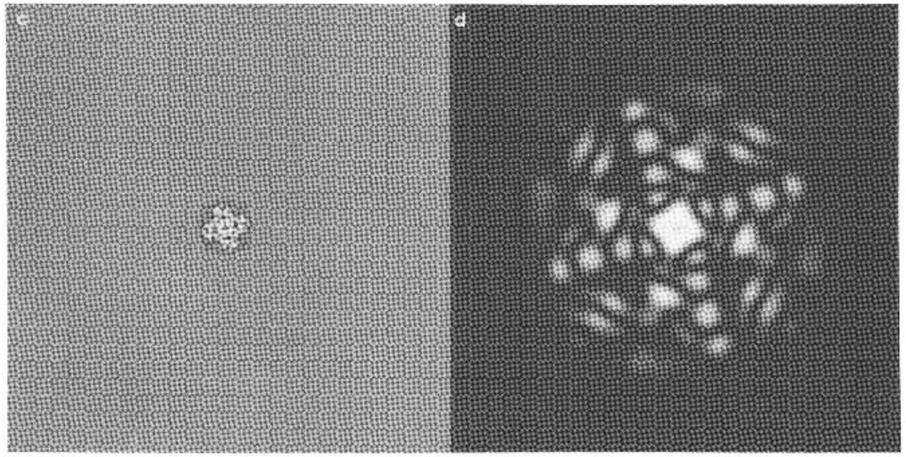
David DeRosier used Markham's lattice to determine how much to shift, and performed averaging by using Adobe Photoshop.

Averaging in 2D crystals

How much and which direction to ship the image can be determined easily from FT of the image of a 2D crystal.







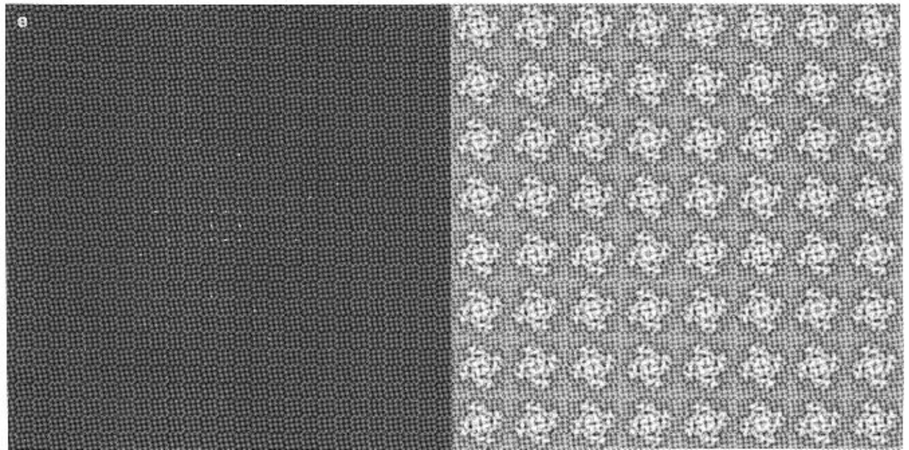


Image averaging in 2D crystal

In 2D crystal, one can extract amplitudes and phases from peaks of FT (contributed by the identical repeats of structural motif) and ignore everything in between peaks (contributed by the random noise). A reverse Fourier Transform using extracted amplitude and phases will give us an averaged features. This is equivalent to the averaging.

It is easy to perform averaging in 2D crystal. The molecules in the 2D crystal are identical in composition and orientation.

What about single molecules

A single particle image data set is a collection of images, each contains projection images of one molecules. The orientations and position of particles in all images are different. Before averaging, one needs to:

- judge how similar is the two particles: *cross-correlation coefficient*;
- shifts/rotates one particle to match another by maximizing ccc: *alignment*;
- separate different particles for averaging: *classification*;

Alignment ⟷ Classification

A digital image is collection of numbers in a grid

3	20	5	-3	4
3	5	34	45	4
0	-2	34	45	6
-1	34	2	3	1
4	5	2	2	0

$$f = \sum_{j=1}^{J} f(\vec{r}_j) = \sum_{j=1}^{J} f(m_j, n_j)$$

3	2	5	-3	4
25	2	4	2	4
0	34	45	5	6
-1	32	40	2	1
35	3	2	2	0

$$g = \sum_{j=1}^{J} g(\vec{r}_j) = \sum_{j=1}^{J} g(m_j, n_j)$$

Cross-correlation coefficient

Cross-correlation coefficient is a measure of similarity and statistical interdependence between two data sets. The mathematic definition of cross-correlation coefficient is:

$$\rho = \frac{\sum_{j=1}^{J}\left[f_1(\vec{r}_j) - \langle f_1 \rangle \right]\left[f_2(\vec{r}_j) - \langle f_2 \rangle \right]}{\left\{ \sum_{j=1}^{J}\left[f_1(\vec{r}_j) - \langle f_1 \rangle \right]^2 \sum_{j=1}^{J}\left[f_2(\vec{r}_j) - \langle f_2 \rangle \right]^2 \right\}^{1/2}}$$

Where:

$$\langle f_i \rangle = \frac{1}{J}\sum_{j=1}^{J} f_i(\vec{r}_j)$$

Note that:

$$-1 < \rho < 1$$

Alignment between two images

Alignment is a process to search the grids to maximize the cross-correlation coefficient between two images. Three parameters are used to define alignment of 2D images: in-plane shift (x,y) and in-plane rotation angle.

Cross-correlation function based alignment:

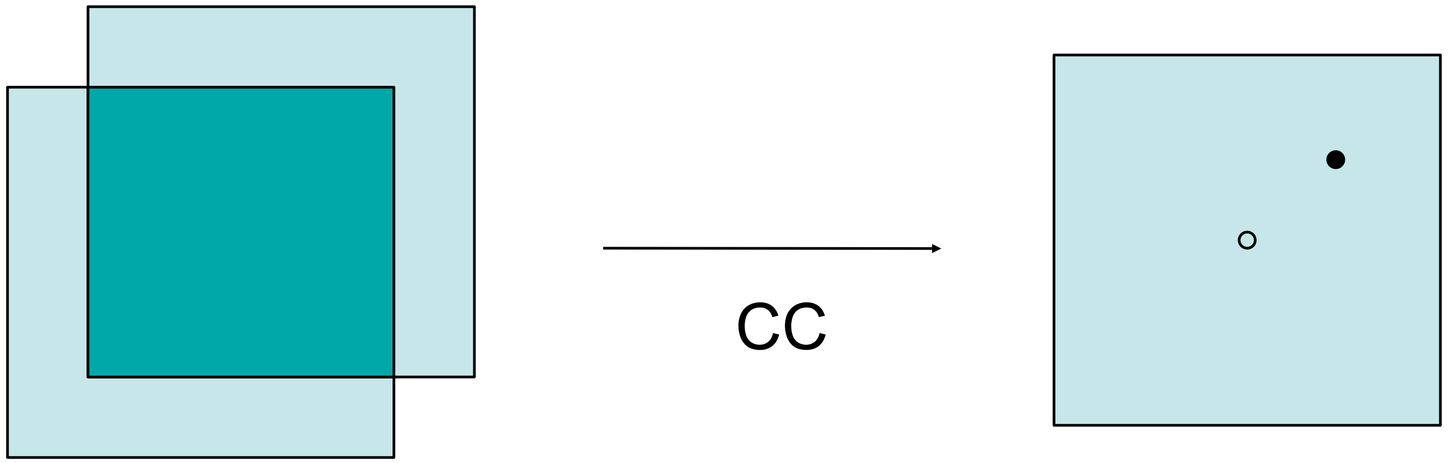
- In-plane shift can be determined by determine the peak position in the translational cross-correlation function between two images.
- Rotation can be determined by different ways: rotational cross-correlation function, Radon transform.

Cross-correlation function

The cross-correlation function is the most important tool for alignment of two images.

The mathematic definition of cross-correlation is:

$$f * g = \int_{-\infty}^{\infty} f(t)g(t-\tau)d\tau$$



Q: what happens if shift is more than half of the image size?

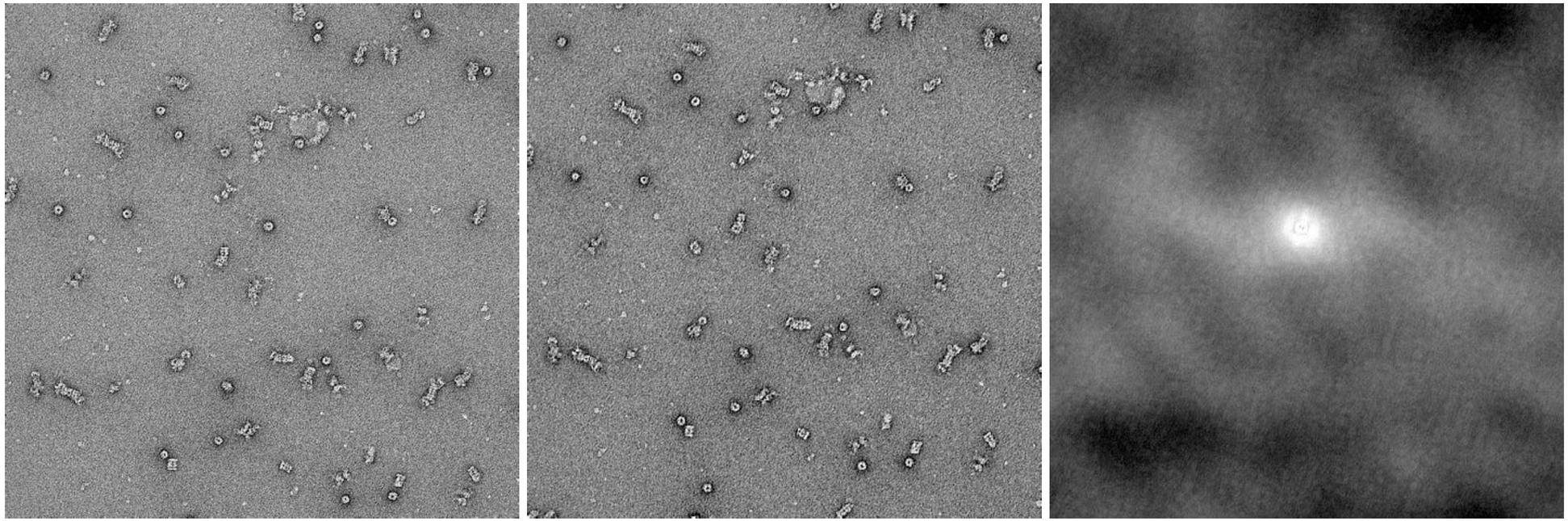
Calculating the cross-correlation

Cross-correlation theorem:

$$f * g = \int_{-\infty}^{\infty} f(t) g(t-\tau) d\tau = F\left\{ F(f) \cdot F^{-1}(g) \right\}$$

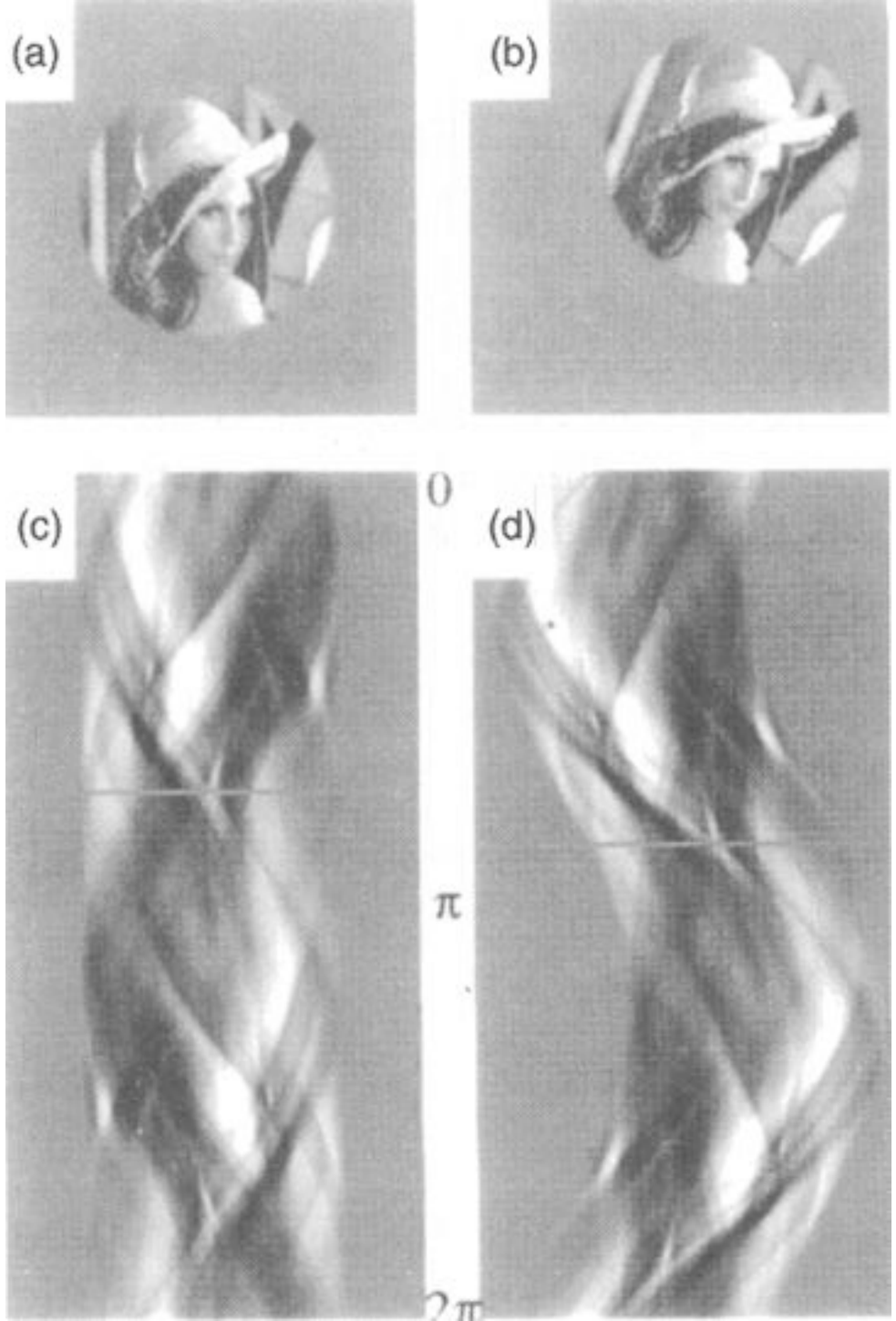
This formula enable us to calculate the cross-correlation between two images easily.

How cross-correlation looks like



-1μm -1.5μm CCF

The image size is 1024X1024. The peak in the CCF is at (445,500). How much is the shift?



Radon transform

Radon transform is an efficient way for determining angular relationship between two images, but it only works well in images with high SNR.

More about the cross-correlation function

• Peak searching in the cross-correlation function;
search for a peak is not just finding the point of highest value in the CCF.

• Keep in mind that one can calculate cross correlation between any two images, and will always find a point with highest value.

• Cross-correlation based alignment and averaging always enhance the features of the reference image.

Demonstration of reference induced bias

Note: The averaged image after reference based alignment is strongly biased towards the reference.

100 images 1000 images reference

From Niko Gorigorieff

Influence of CTF on image

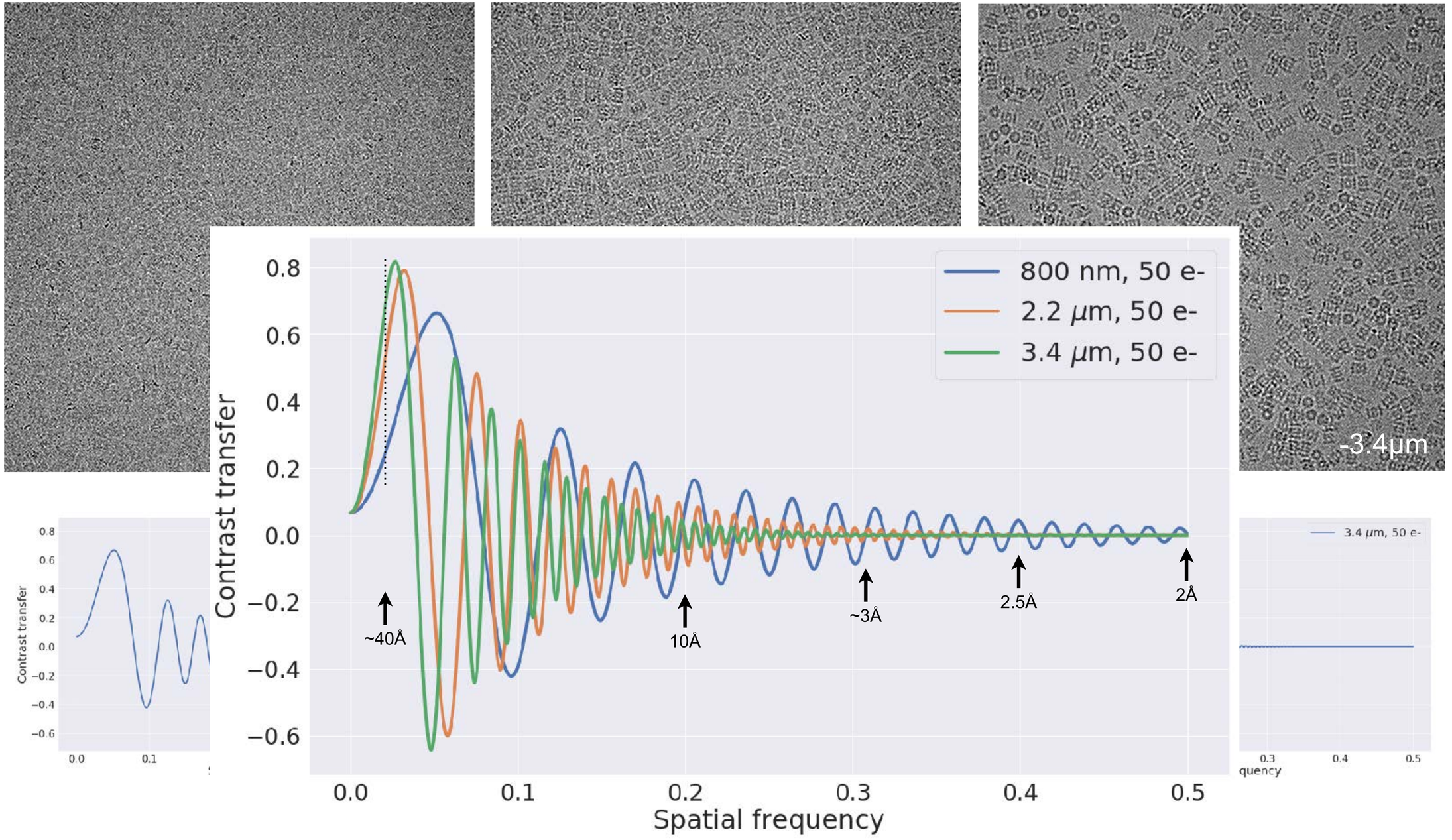
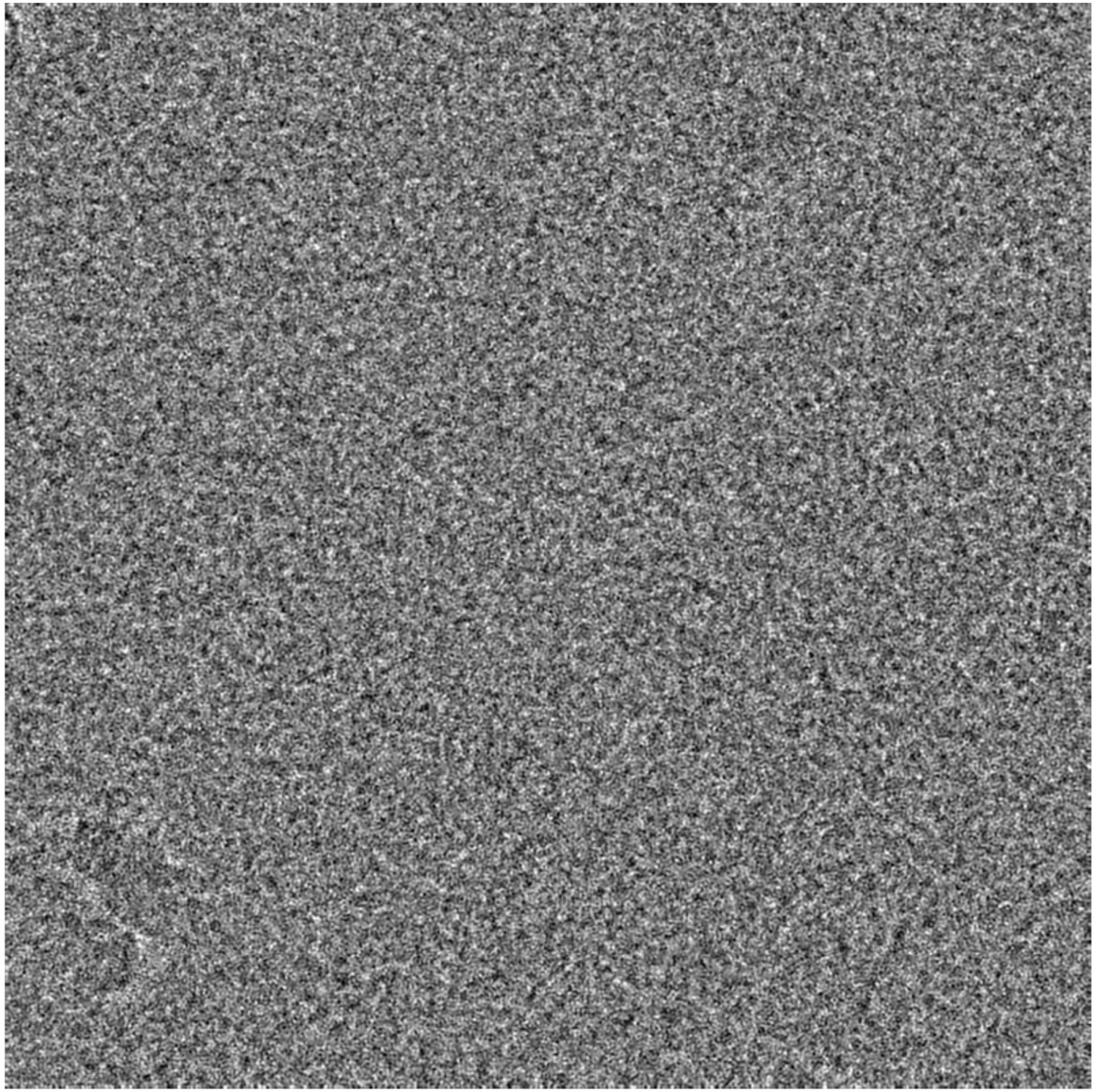
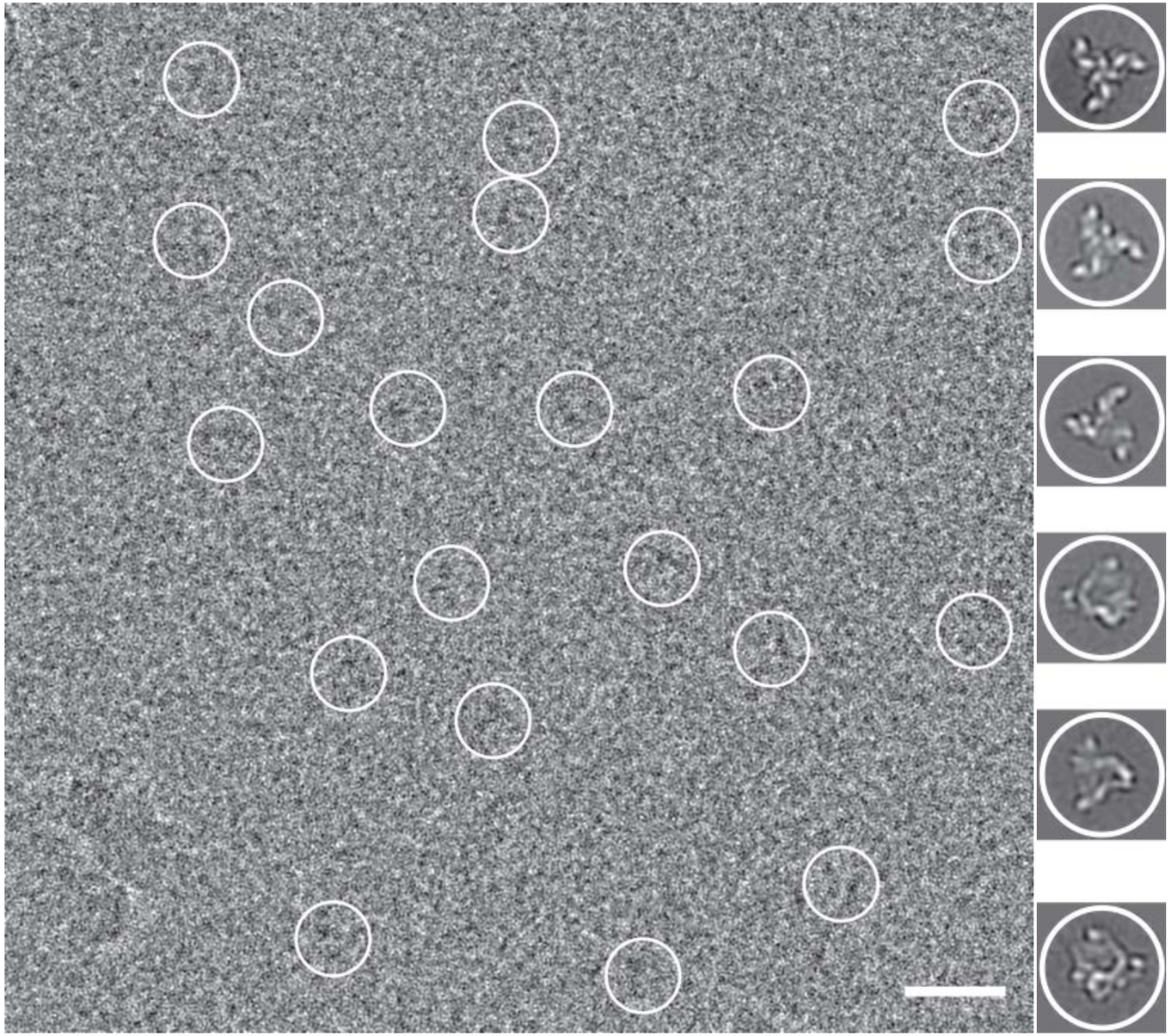


Image taken at close to focus



Mao et al, (2012) NSMB

Image taken at close to focus



Mao et al, (2012) NSMB

Ghost structure of HIV trimer

PNAS

Molecular architecture of the uncleaved HIV-1 envelope glycoprotein trimer

Youdong Mao[a,b,1], Liping Wang[a,b], Christopher Gu[a,b], Alon Herschhorn[a,b], Anik Désormeaux[c], Andrés Finzi[c], Shi-Hua Xiang[d], and Joseph G. Sodroski[a,b,e,f,1]

[a]Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, Boston, MA 02215; [b]Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115; [c]Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Department of Microbiology and Immunology, Université de Montréal, Montréal, QC, Canada H3A 2B4; [d]Nebraska Center for Virology, School of Veterinary Medicine and Biomedical Sciences, University of Nebraska-Lincoln, Lincoln, NE 68583; [e]Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology, and Harvard, Cambridge, MA 02139; and [f]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02115

PERSPECTIVE

Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise

Richard Henderson[1]
Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 0QH, United Kingdom

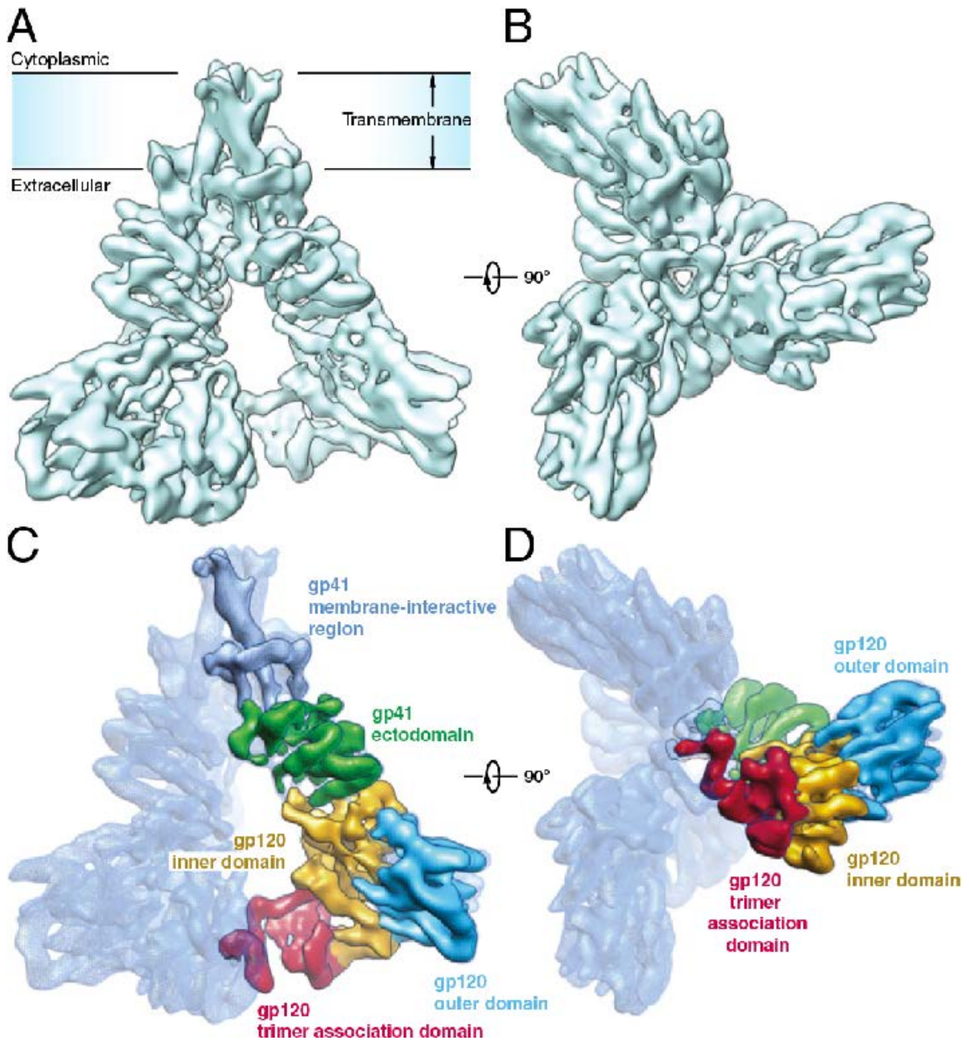
PERSPECTIVE

Finding trimeric HIV-1 envelope glycoproteins in random noise

Marin van Heel[1]
Leiden Institute of Chemistry, Leiden University, 2333 CC Leiden, The Netherlands

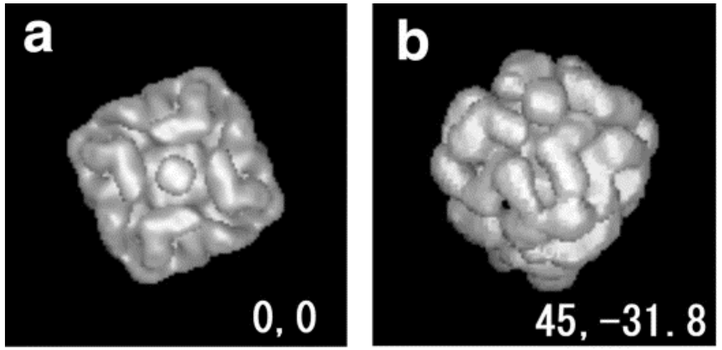
Structure of trimeric HIV-1 envelope glycoproteins

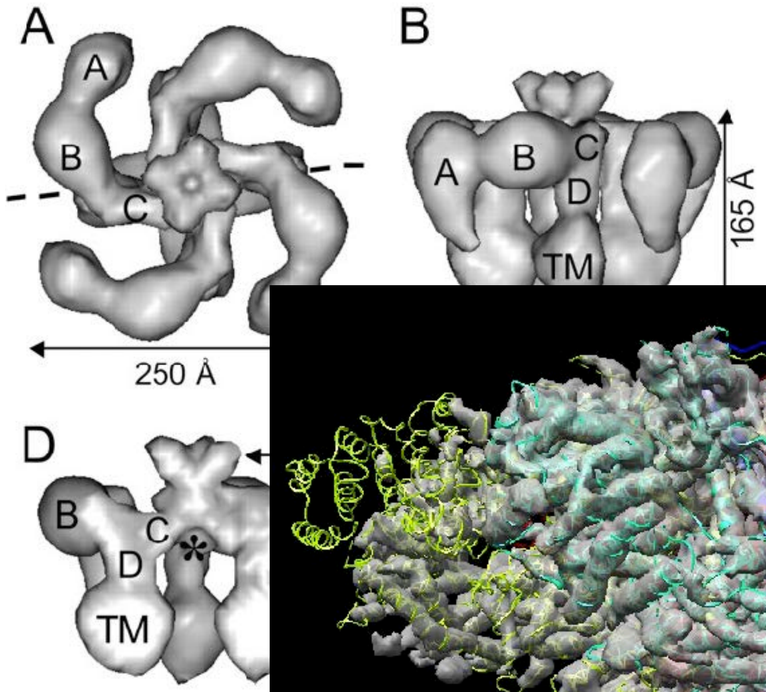
Sriram Subramaniam[1]
Laboratory of Cell Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892

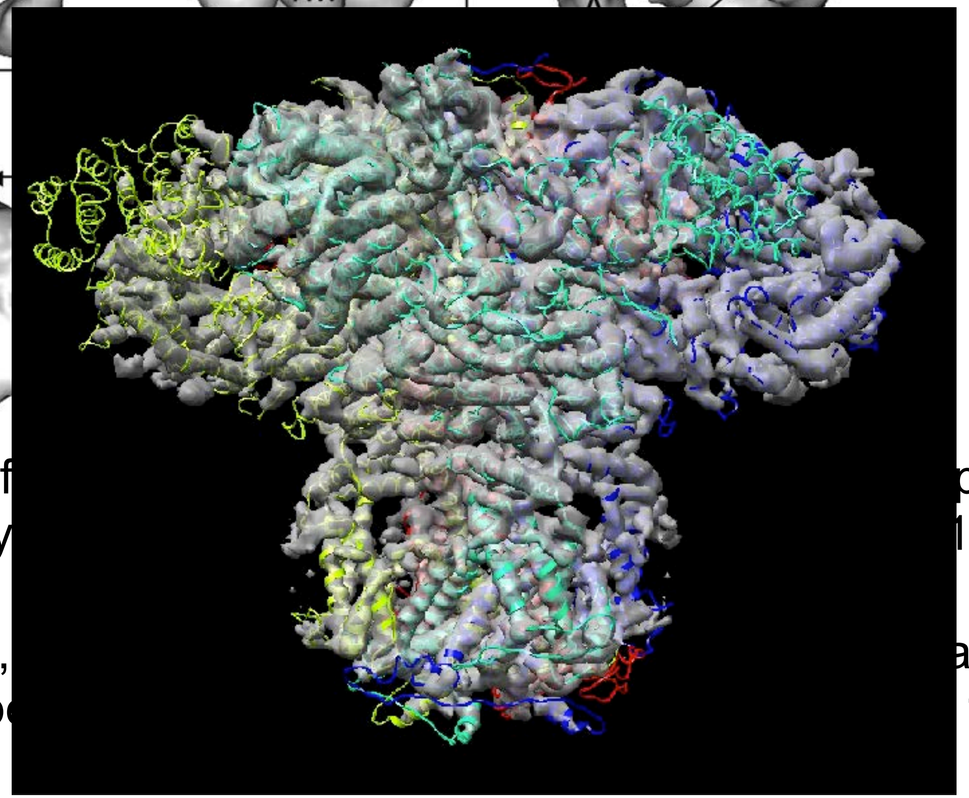


* Cryo-EM is not a turn-key technology and it is possible to make massive mistakes!

Single particle cryo-EM of membrane proteins







Structure of
revealed by

otor
1319-22.

Insoitol 1,4,

avities
155-64.

4.7Å resolution, Nature 2015

Multi-reference alignment

For a heterogeneous data set, multiple references are used. Each images are aligned again each references, and decide which one yields highest cross-correlation coefficient.

Classification

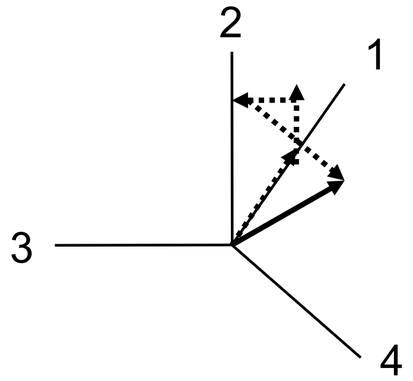
Classification - a process of dividing a set of images into subsets with similar features.

One can perform classification based on CCC to determine if the images are similar with each other; But for a very large data set of very noisy images (> 50,000 images)?

Hyperspace

An image of m×m pixels can be represented by a vector (or end point of a vector) in the hyperspace of m×m dimensions.

3	2
2	3



$$f = (f_1, f_2, \ldots\ldots, f_m) = \sum_{i=1}^{m} f_i \vec{a}_i$$ Where: $$|\vec{a}_i| = 1;$$

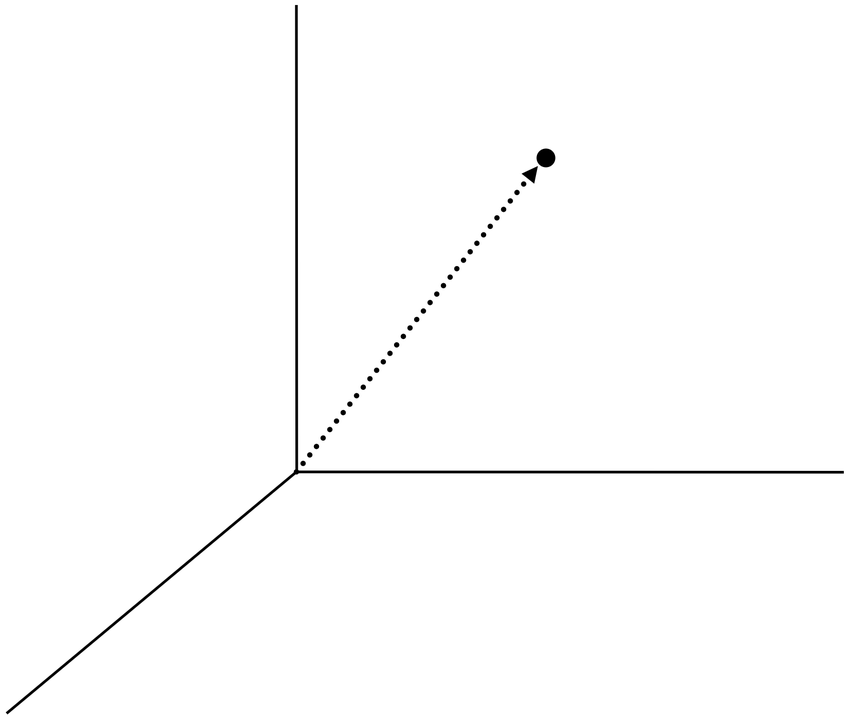
$$\vec{a}_i \perp \vec{a}_j (j \neq i; j = 1, \ldots m);$$

Similar to the cross-correlation coefficient, the distance between two spots in the hyperspace represents the difference between two images.

A data set is represented as a cloud in the hyperspace. The center of the cloud is the average of the all images in the data set.

A data set is represented as a cloud in the hyperspace. The center of the cloud is the average of the all images in the data set.

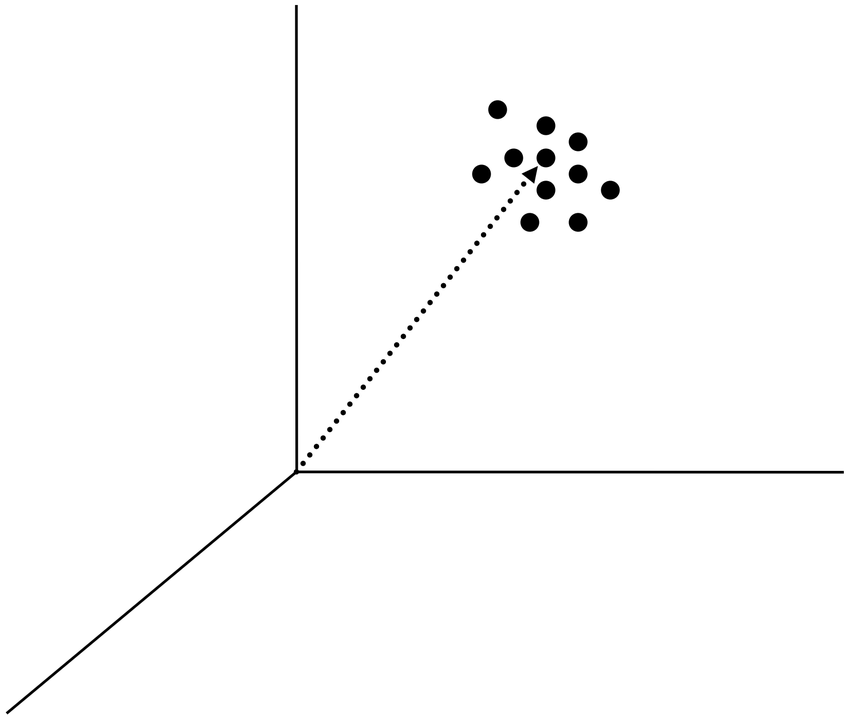
An image without any noise is represented by a point.



A data set is represented as a cloud in the hyperspace. The center of the cloud is the average of the all images in the data set.

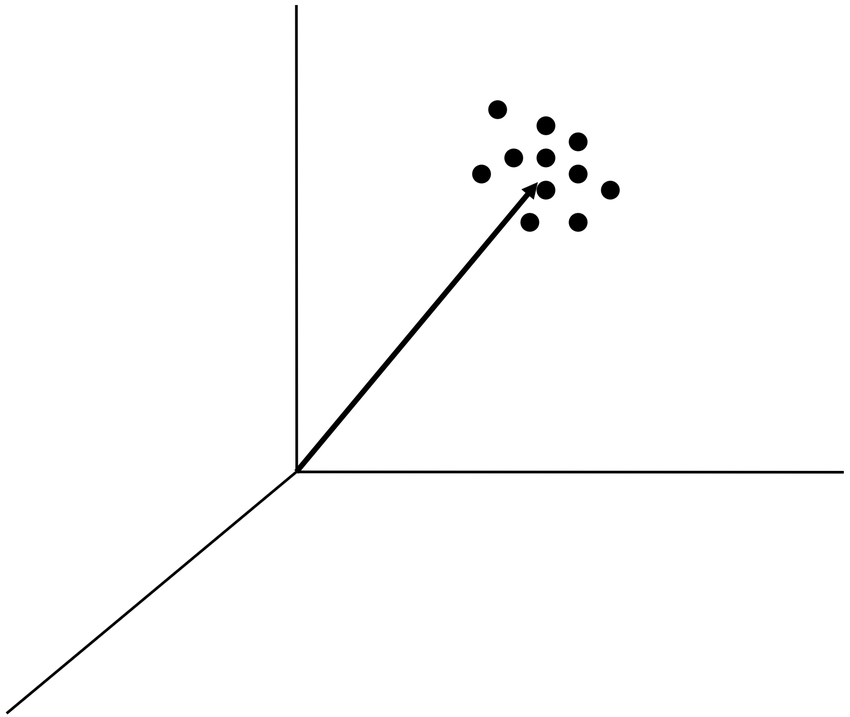
An image without any noise is represented by a point.

Adding random noise to the image expand the point into a cloud.



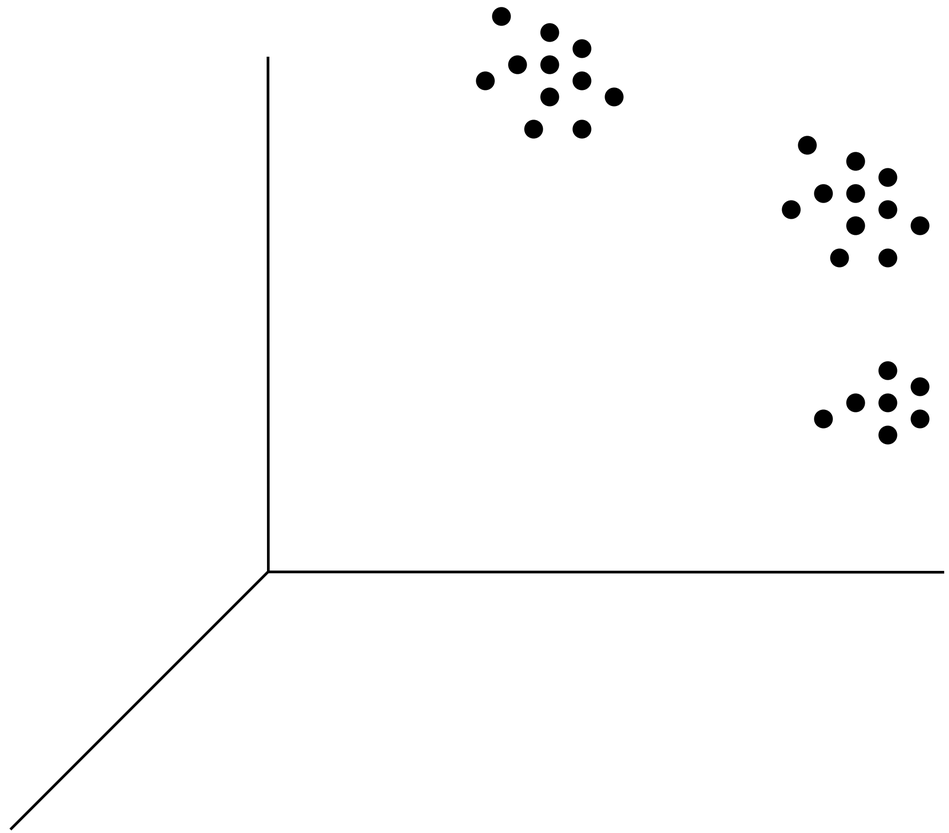
A data set is represented as a cloud in the hyperspace. The center of the cloud is the average of the all images in the data set.

The center of the loud is the average.



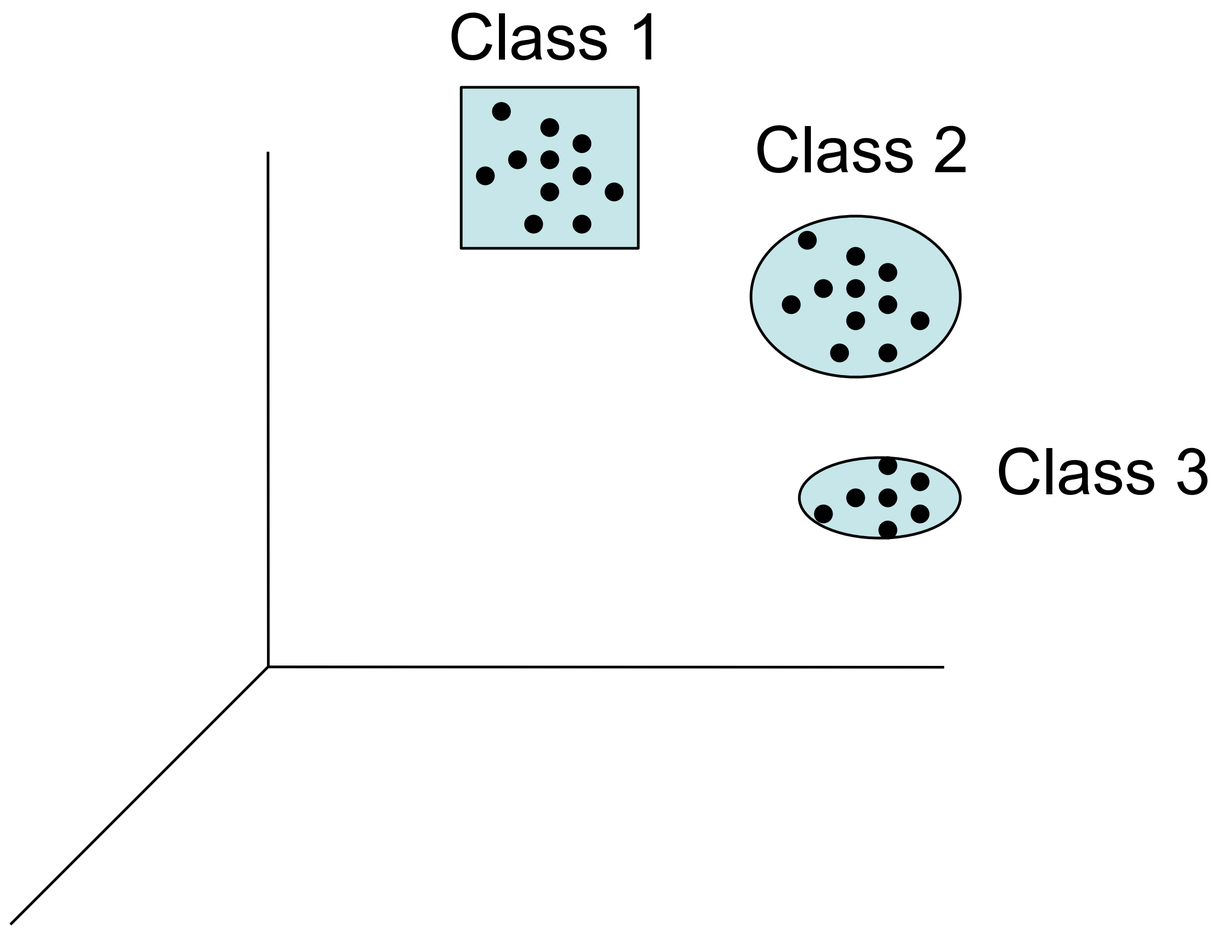
Classification

Assume images are aligned with each other. The clouds of particles can be grouped into different groups - classification.



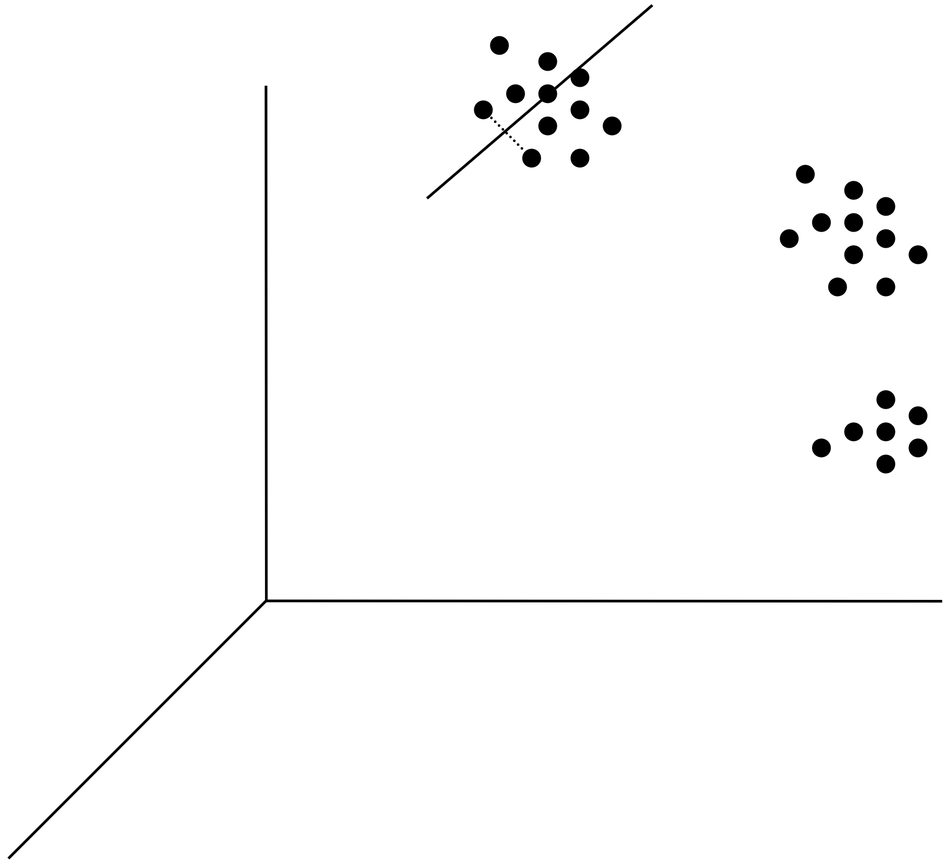
Classification

Assume images are aligned with each other. The clouds of particles can be grouped into different groups - classification.



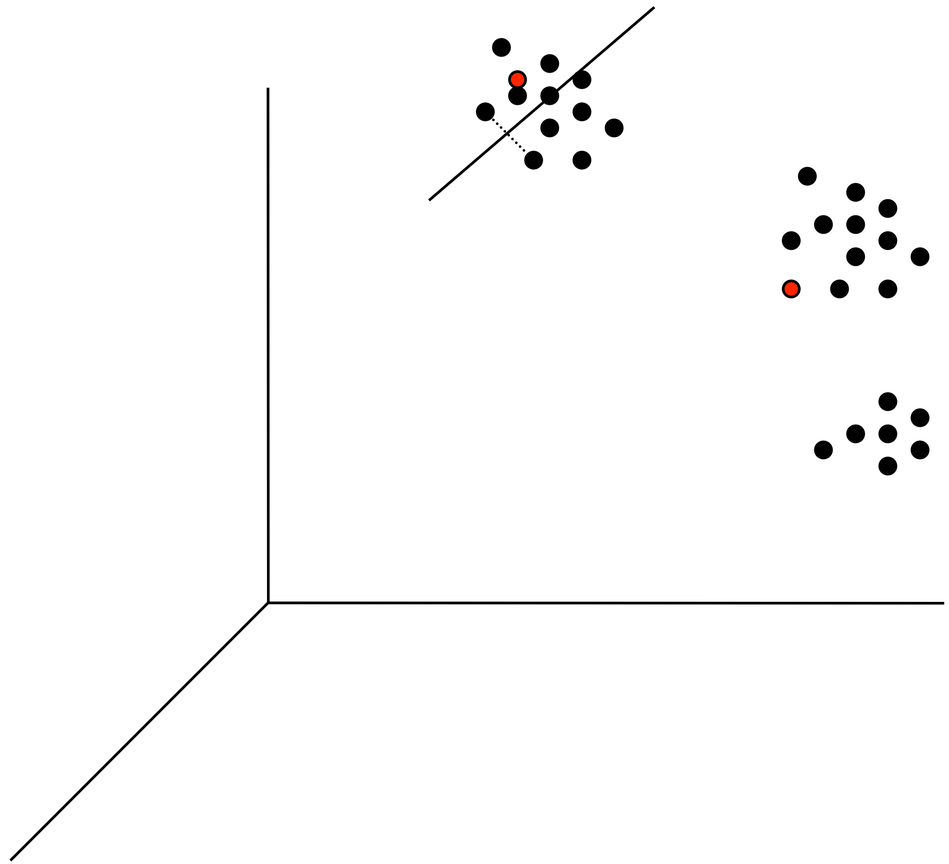
Classification

Assume images are aligned with each other. The clouds of particles can be grouped into different groups - classification.



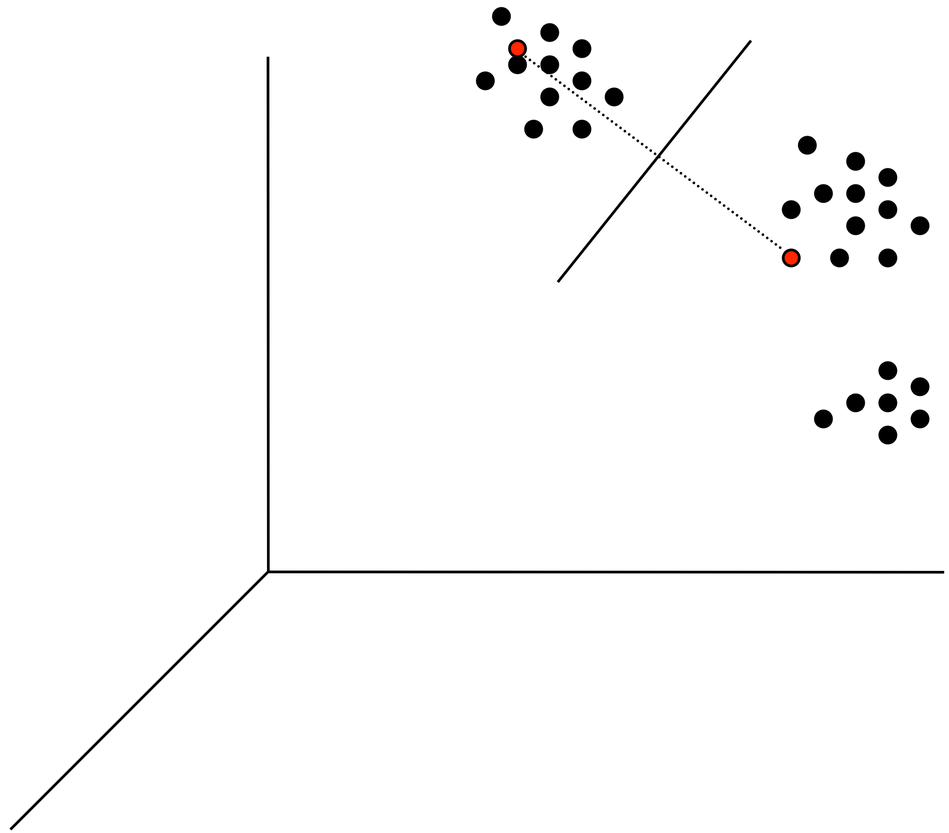
Classification

Assume images are aligned with each other. The clouds of particles can be grouped into different groups - classification.



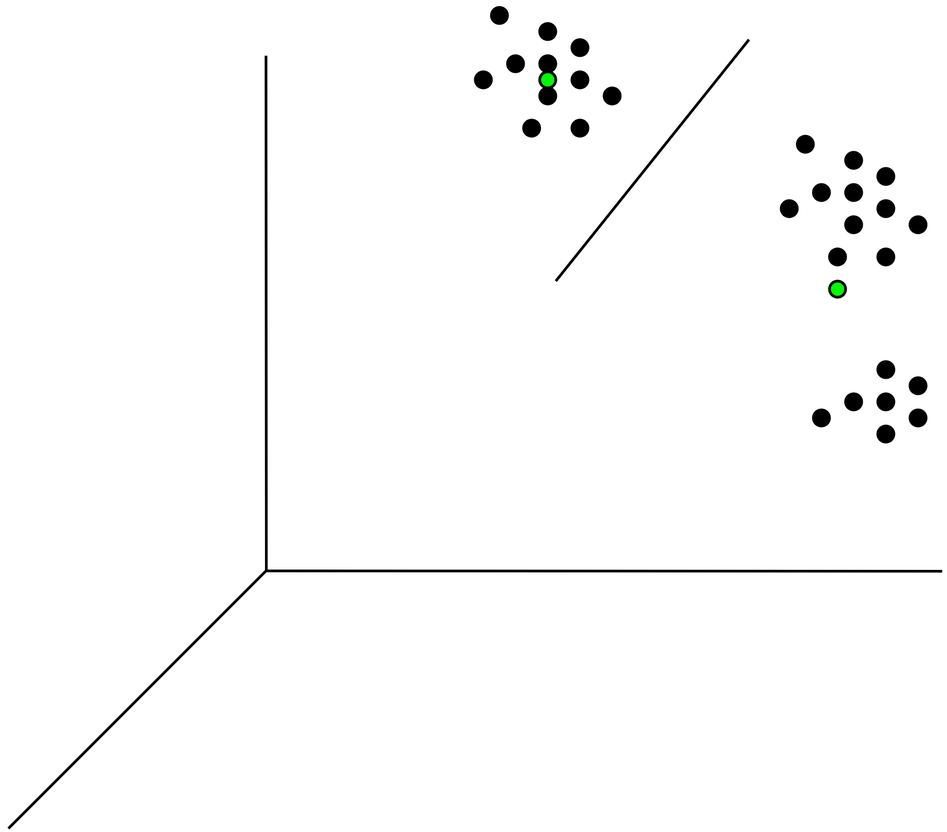
Classification

Assume images are aligned with each other. The clouds of particles can be grouped into different groups - classification.



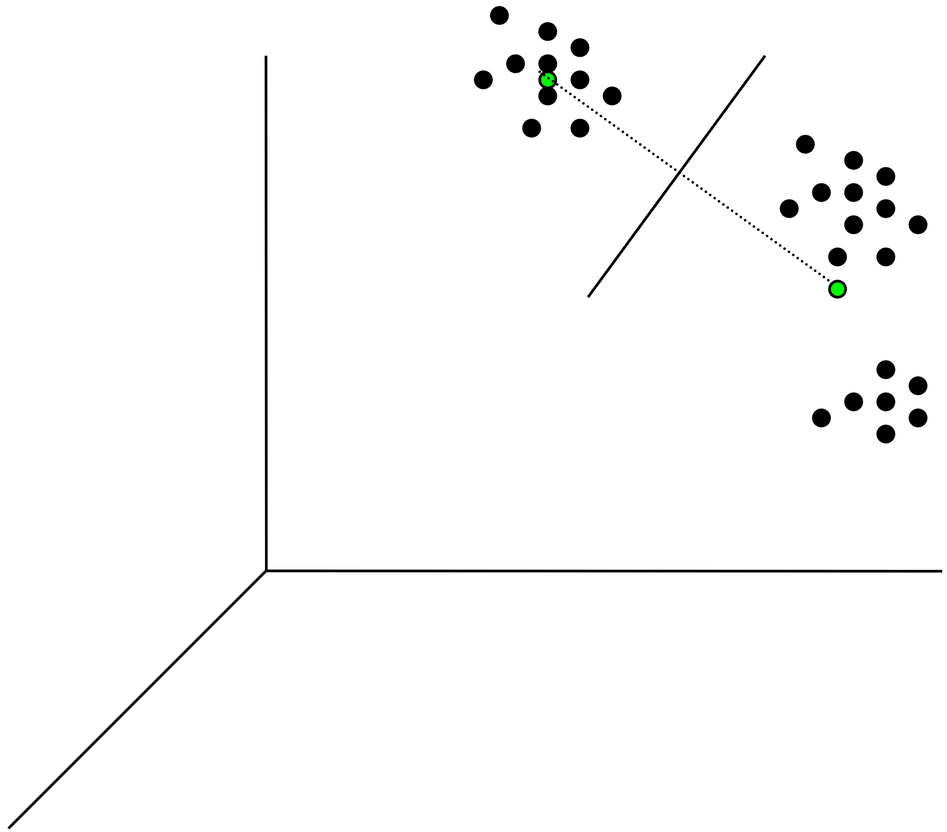
Classification

Assume images are aligned with each other. The clouds of particles can be grouped into different groups - classification.



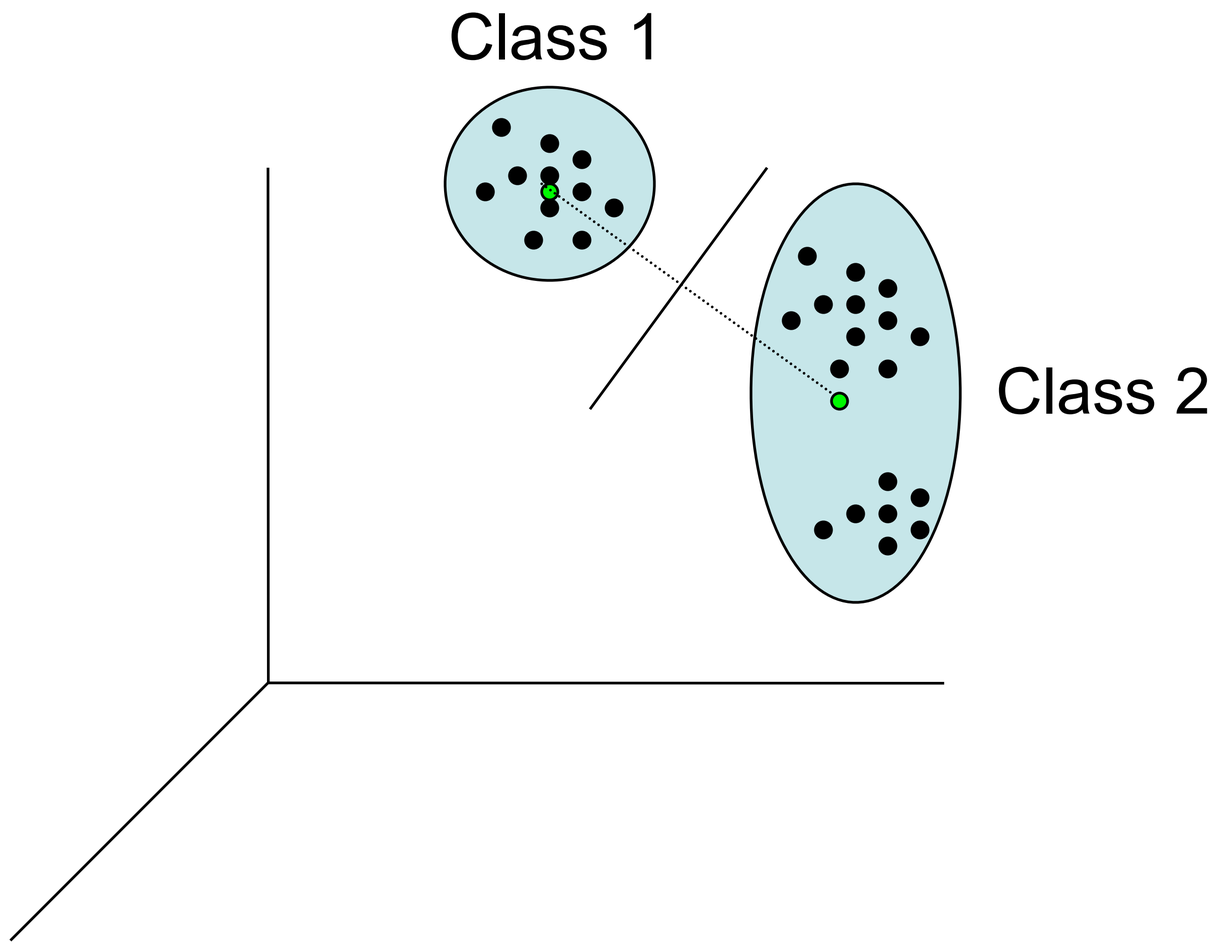
Classification

Assume images are aligned with each other. The clouds of particles can be grouped into different groups - classification.



Classification

Assume images are aligned with each other. The clouds of particles can be grouped into different groups - classification.



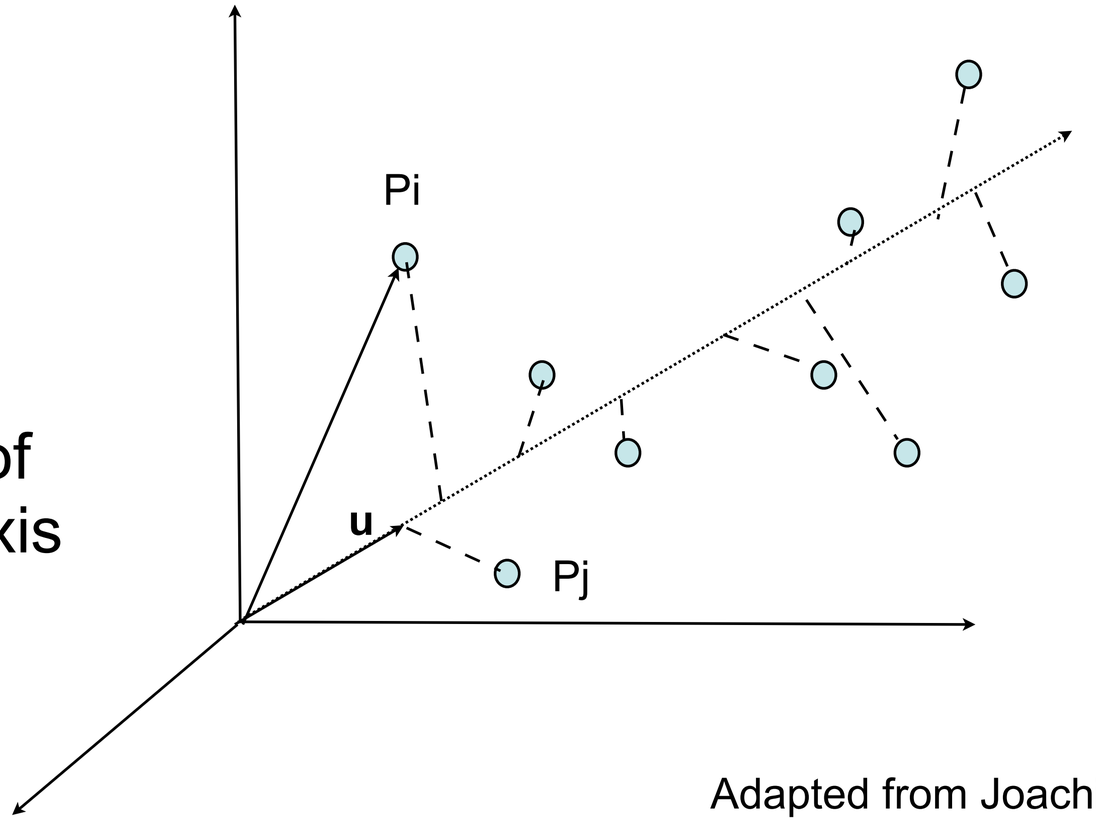
K-mean classification

Multivariate statistical analysis

Making patterns emerge from data

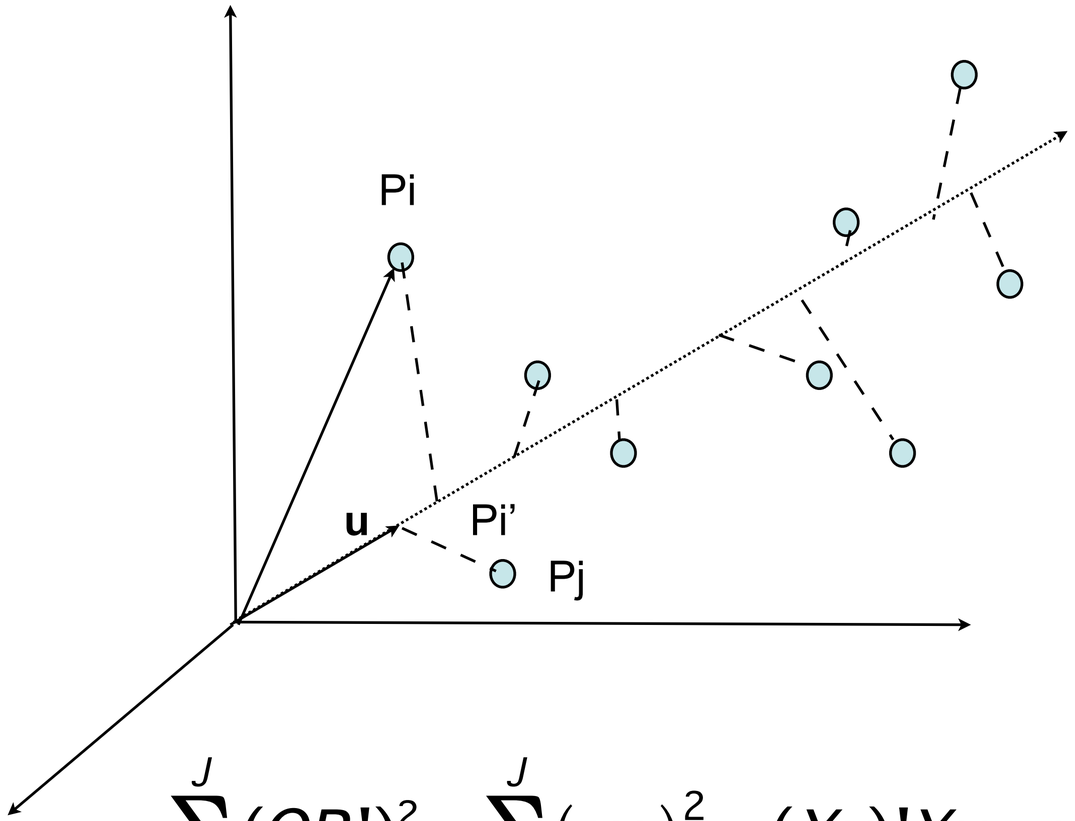
Multivariate statistical analysis:
- Principal Component Analysis
- Correspondence Analysis

Definition of principal axis



Adapted from Joachim Frank

Principal component analysis (PCA)



$$\sum_{j=1}^{J}(OP_i')^2 = \sum_{j=1}^{J}(x_i u)^2 = (Xu)'Xu = u'X'Xu \longrightarrow \max$$

with constraint: u'u = 1 X: coordinate matrix

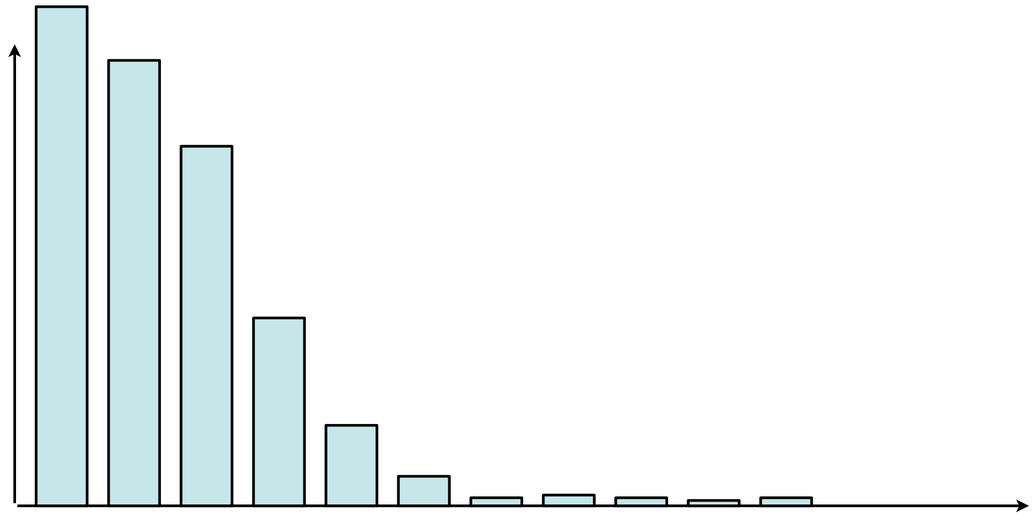
Eigenvector-eigenvalue equation

$$\mathbf{Du} = \lambda\mathbf{u}$$

where $D = (X - \bar{X})'(X - \bar{X})$

Solution of this equation generate a set of eigenvectors and eigenvalues.

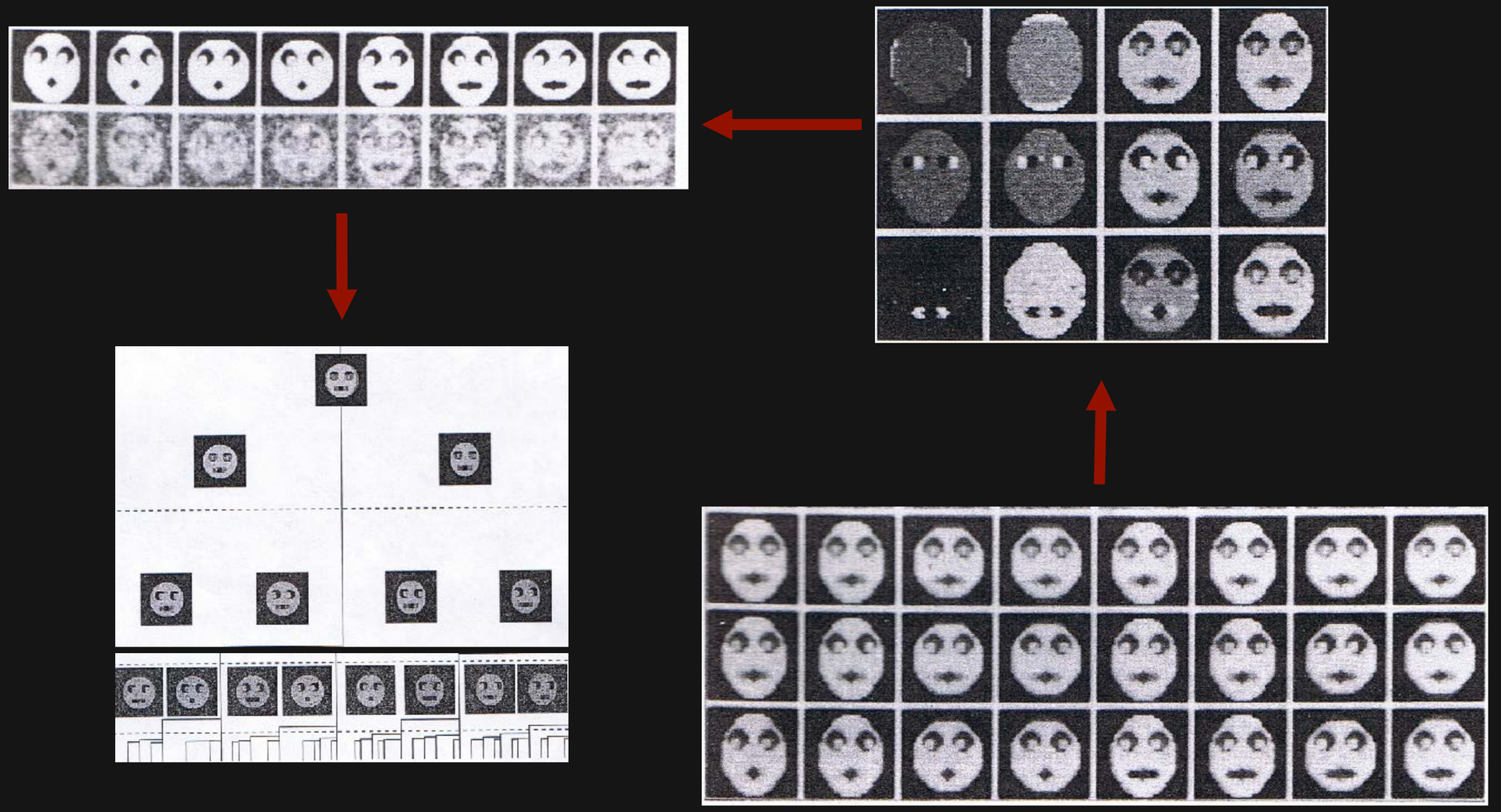
Significant factors:



Classification based on eigenvector/eigenvalue clustering;

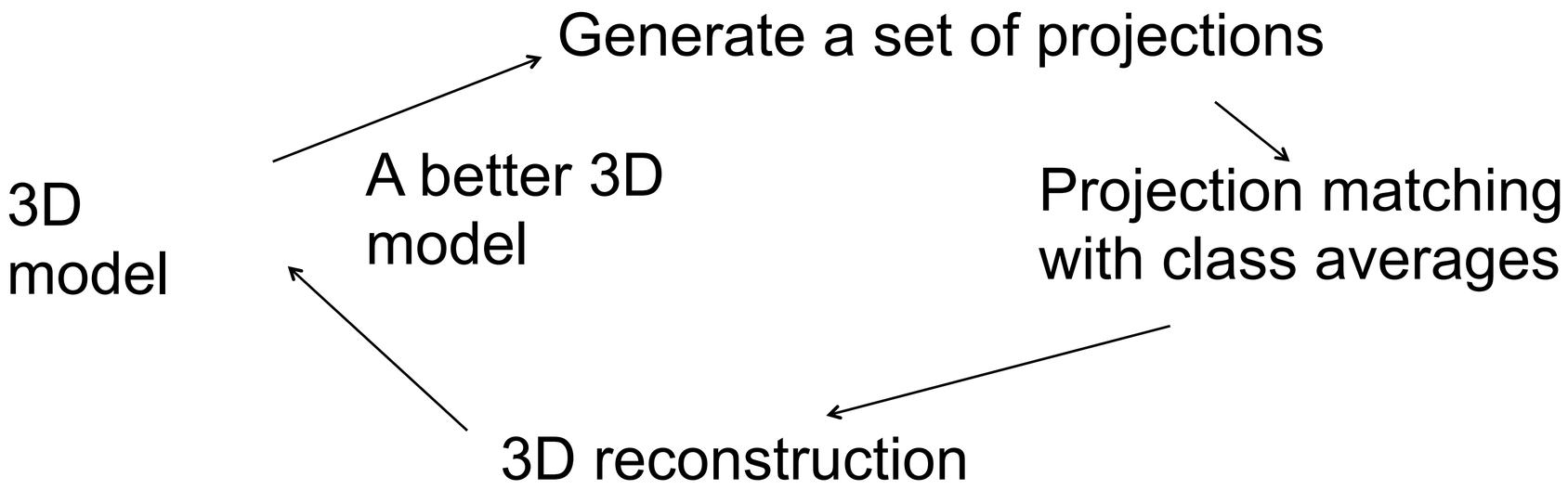
Multivariate statistical analysis & classification of images

Principle Component Analysis



Iterative refinement procedure

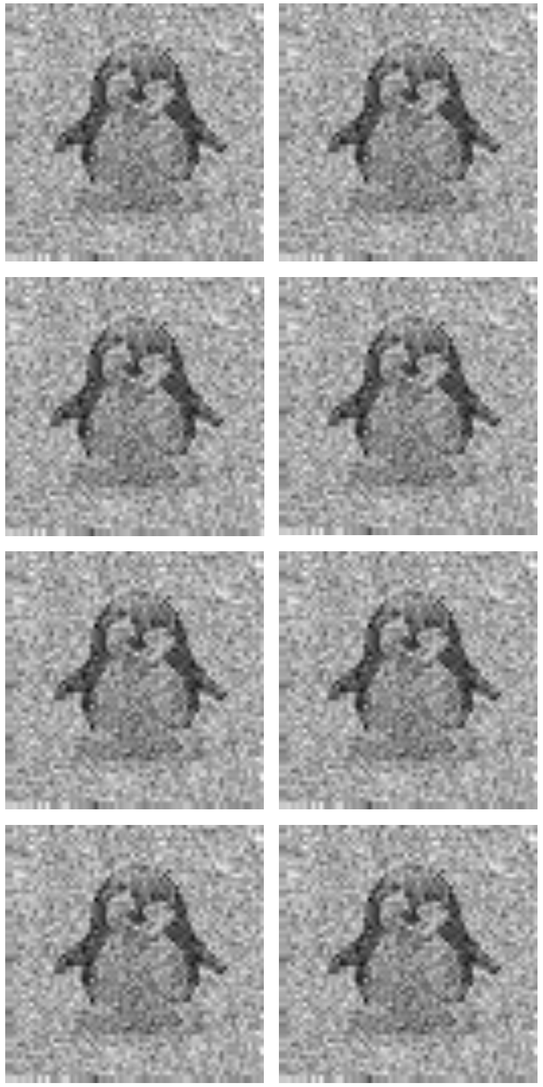
Iterative refinement procedure, using reference model based projection matching:



A least square approach to find the best solution that matches all data.

Image averaging

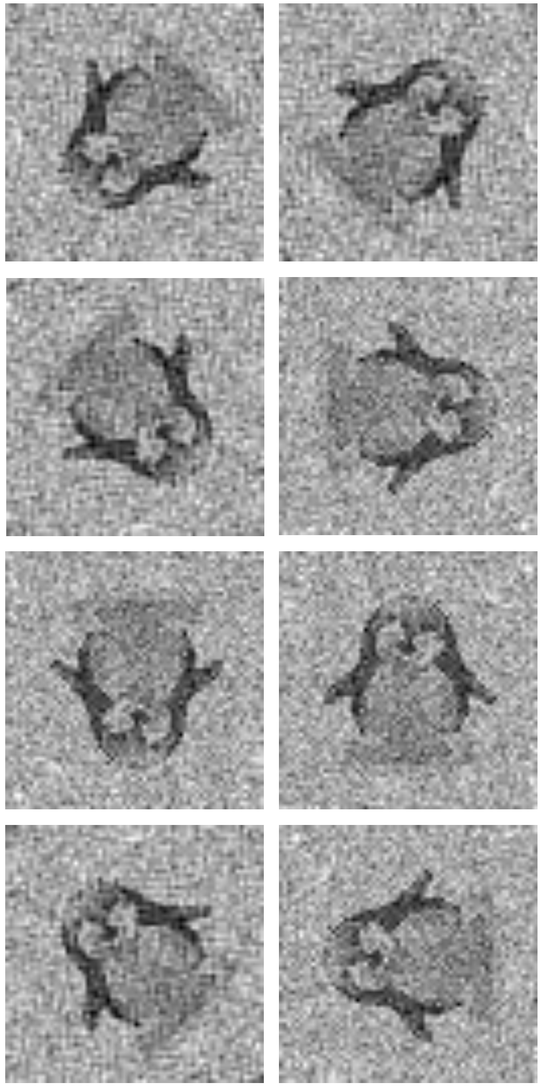
Averaging of a large number of identical images improves the SNR. A complete problem is simple to solve.



$$Average = \frac{1}{N}\sum_{i=1}^{N} X^{i}$$

Observed data (*X*): images

But we have an incomplete data set



Observed data (*X*): images

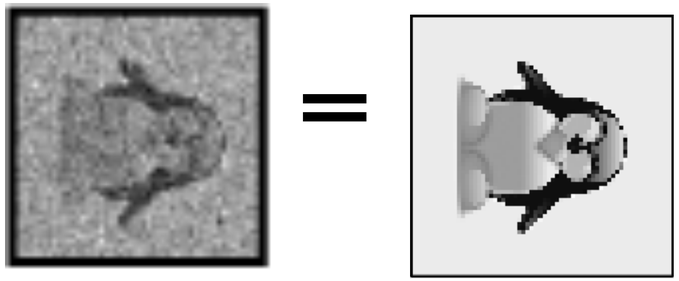
Missing data (Y):

Rotations, translations, classes & conformations

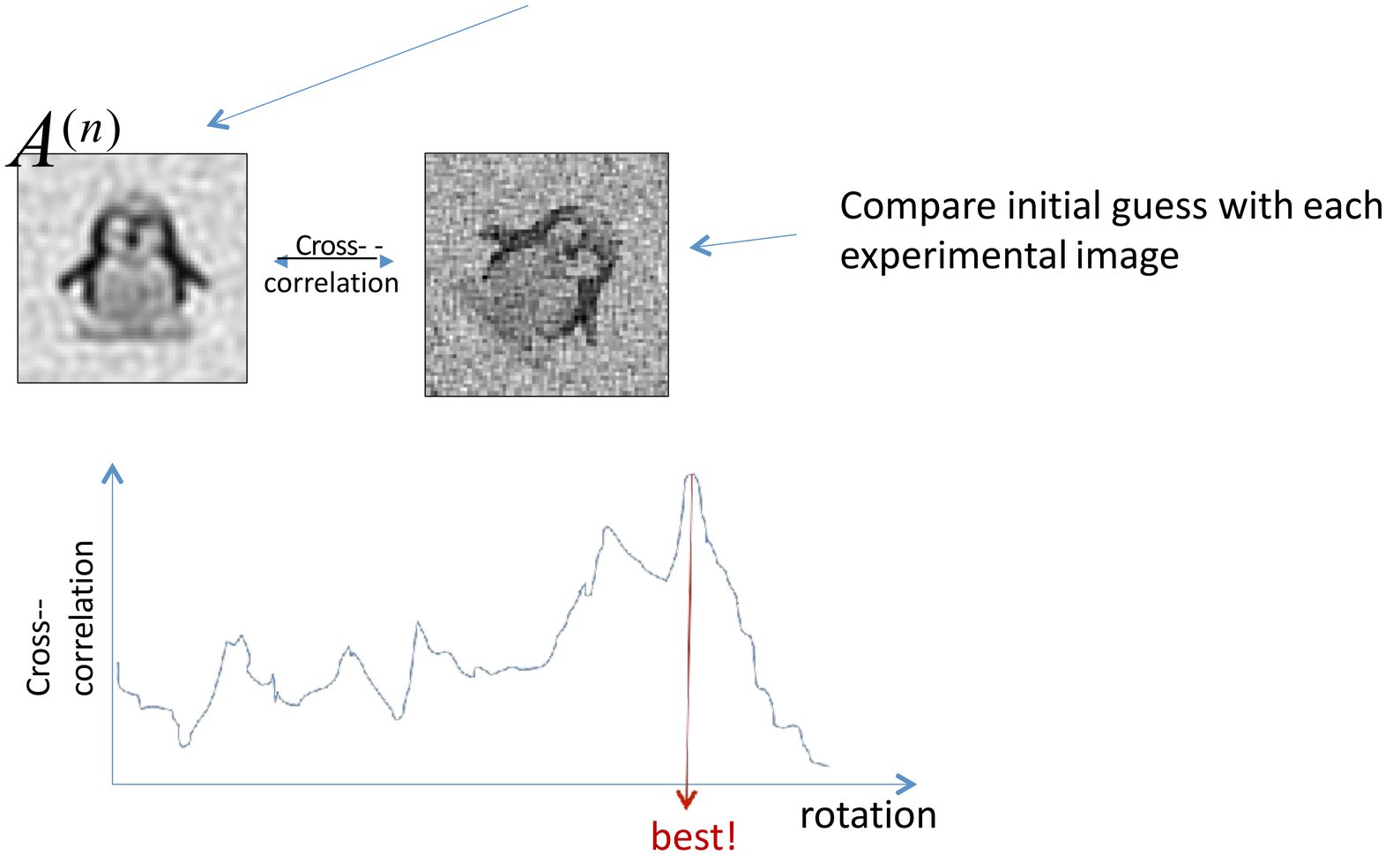
How do we find Y?

$$X_i = P_\varphi V_k$$

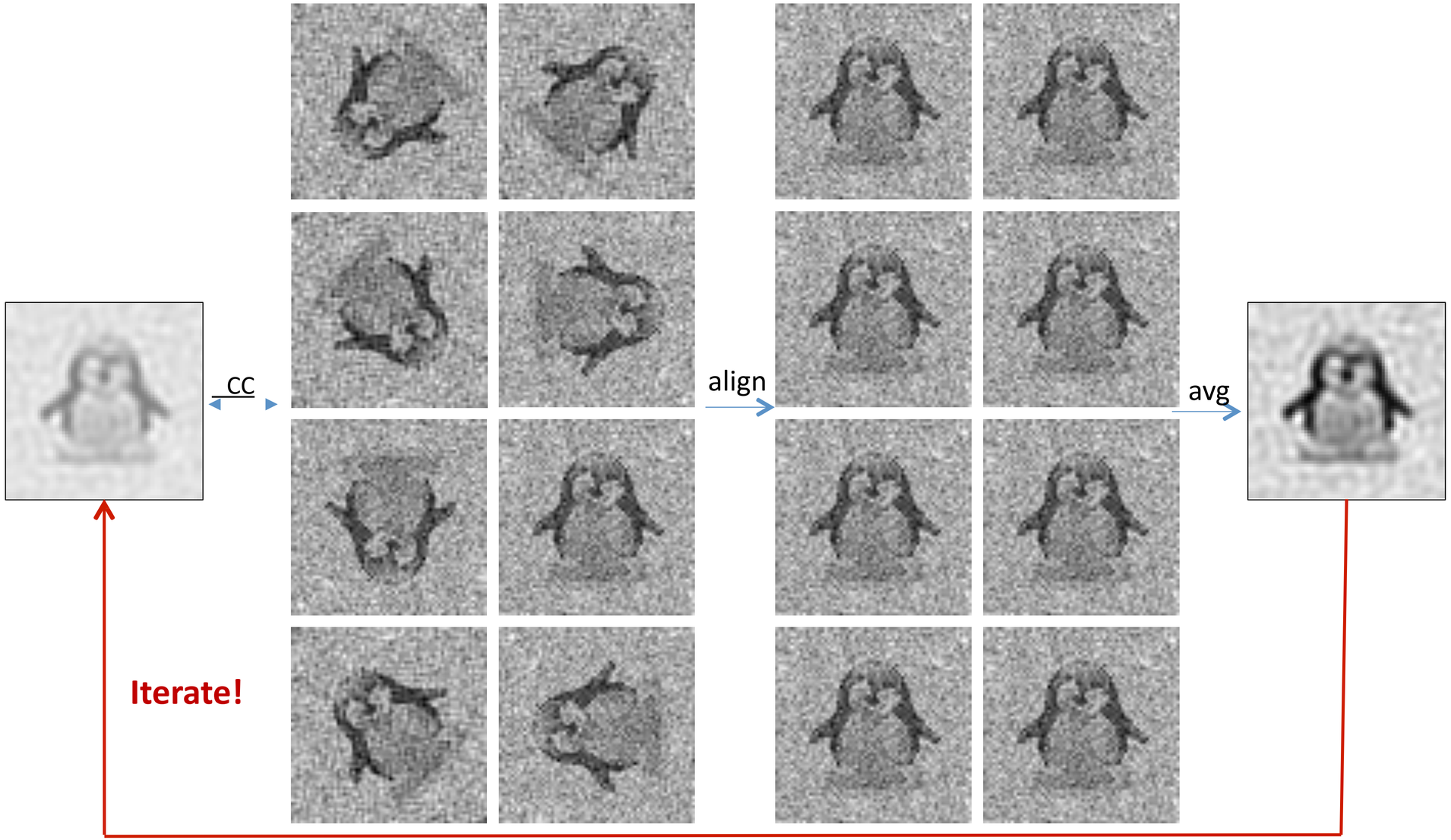
$$X_i\,(\text{Observed Projection}) = P_\varphi\,(\text{Rotations, etc})\,V_k\,(\text{Actual Object})$$



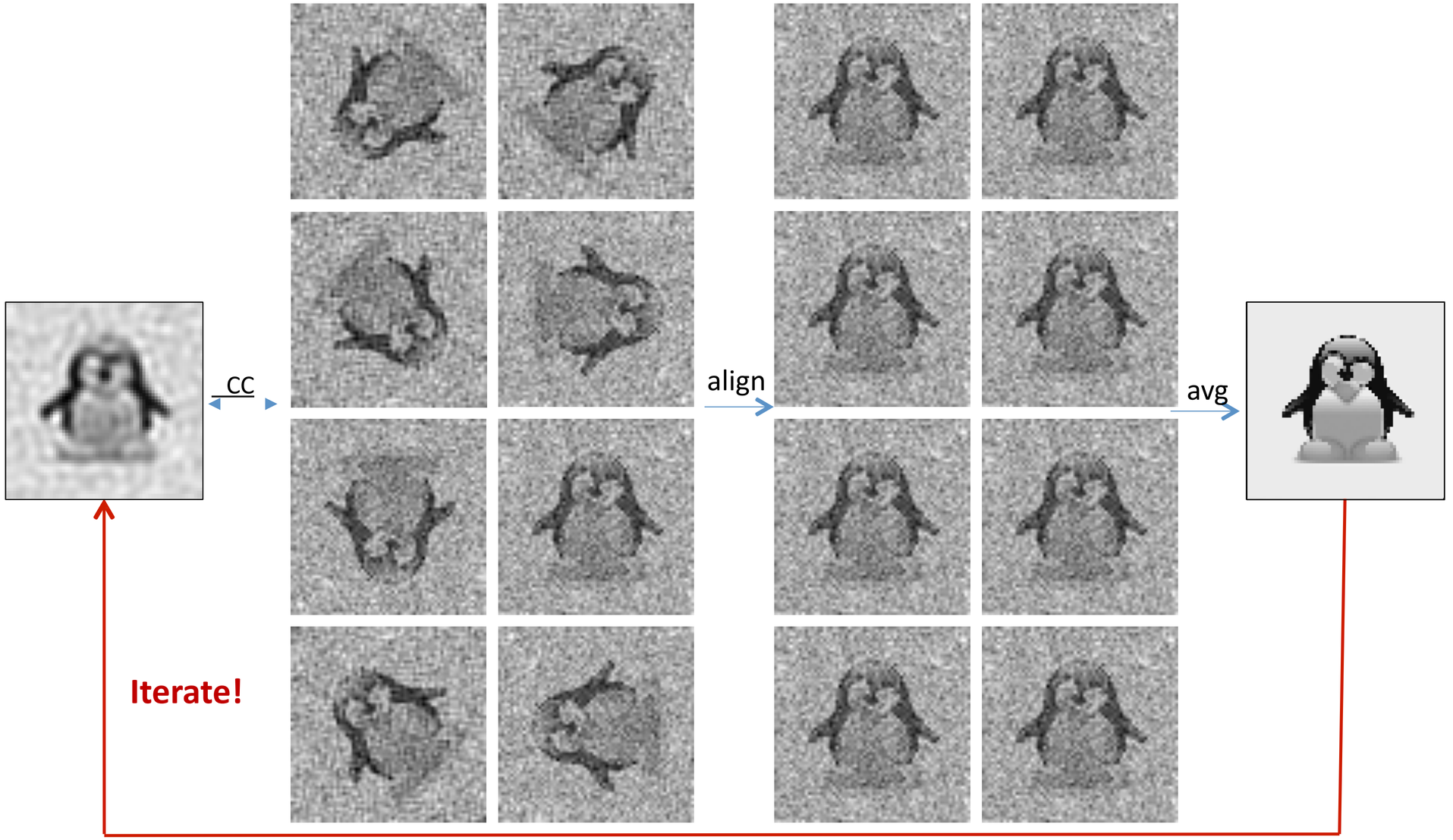
- Starts from some initial **guess** (source of model bias) about the structure



Iteratively align and average
How big is the search space?



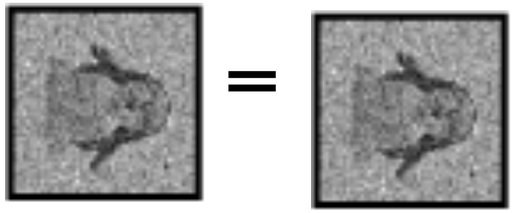
Iteratively align and average
How big is the search space?



1) Uses model projections that include noise

$$X_i = P_\varphi V_k + N_i$$

$$X_i\,(\text{Observed Projection}) = P_\varphi\,(\text{Rotations, etc})\,V_k\,(\text{Actual Object}) + \text{Noise}$$



Sjors Scheres

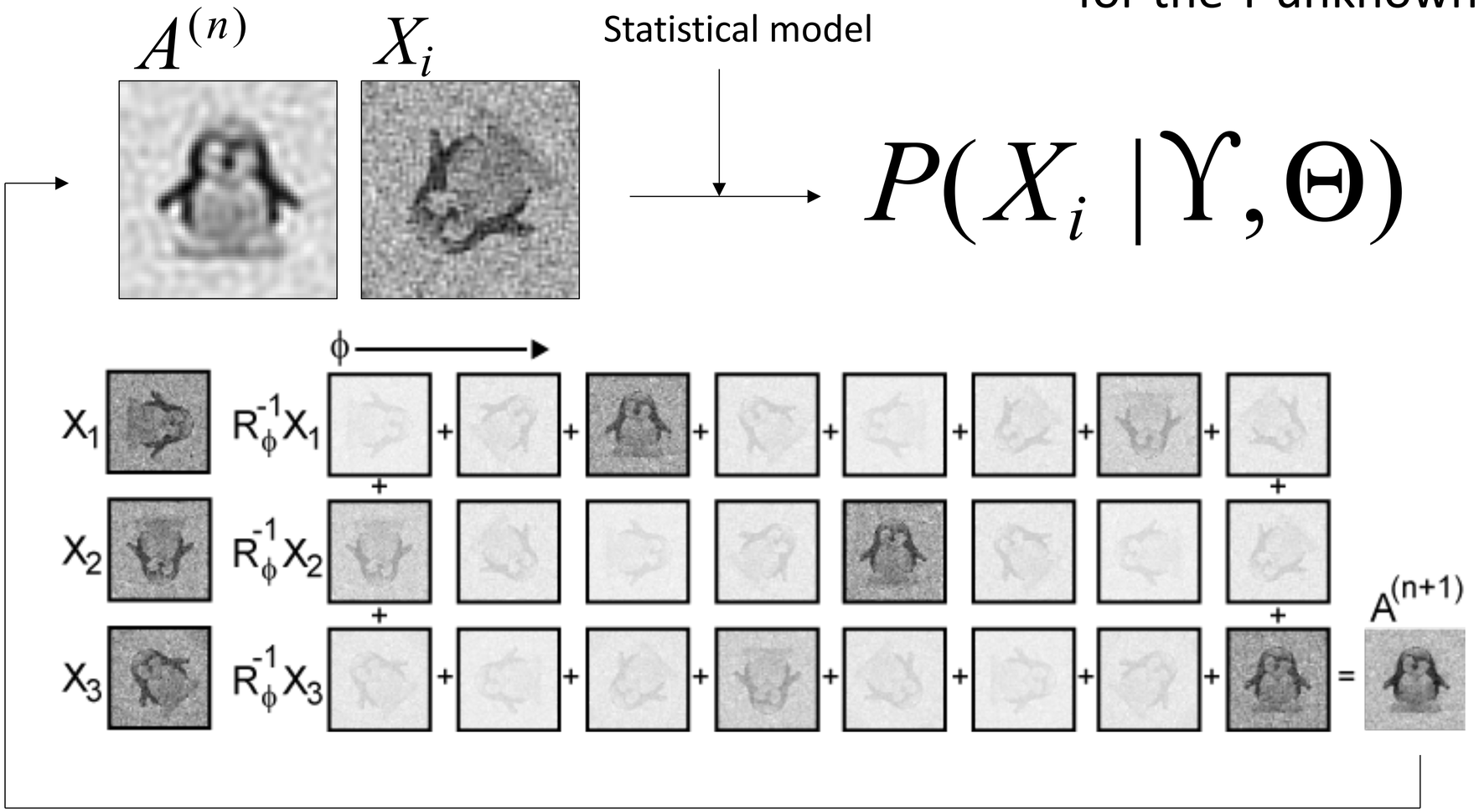
2) Maximizes likelihood with “marginalization” over Y

Need more? See *Methods in Enzymology*, **482** (2010)

Does not assign discrete “best” values for the Y unknowns !

Iteratively maximize the likelihood of observing a given image (X_i), given the model (Θ) + and the values of the unknown parameters (Y).

Θ and Y change each cycle.

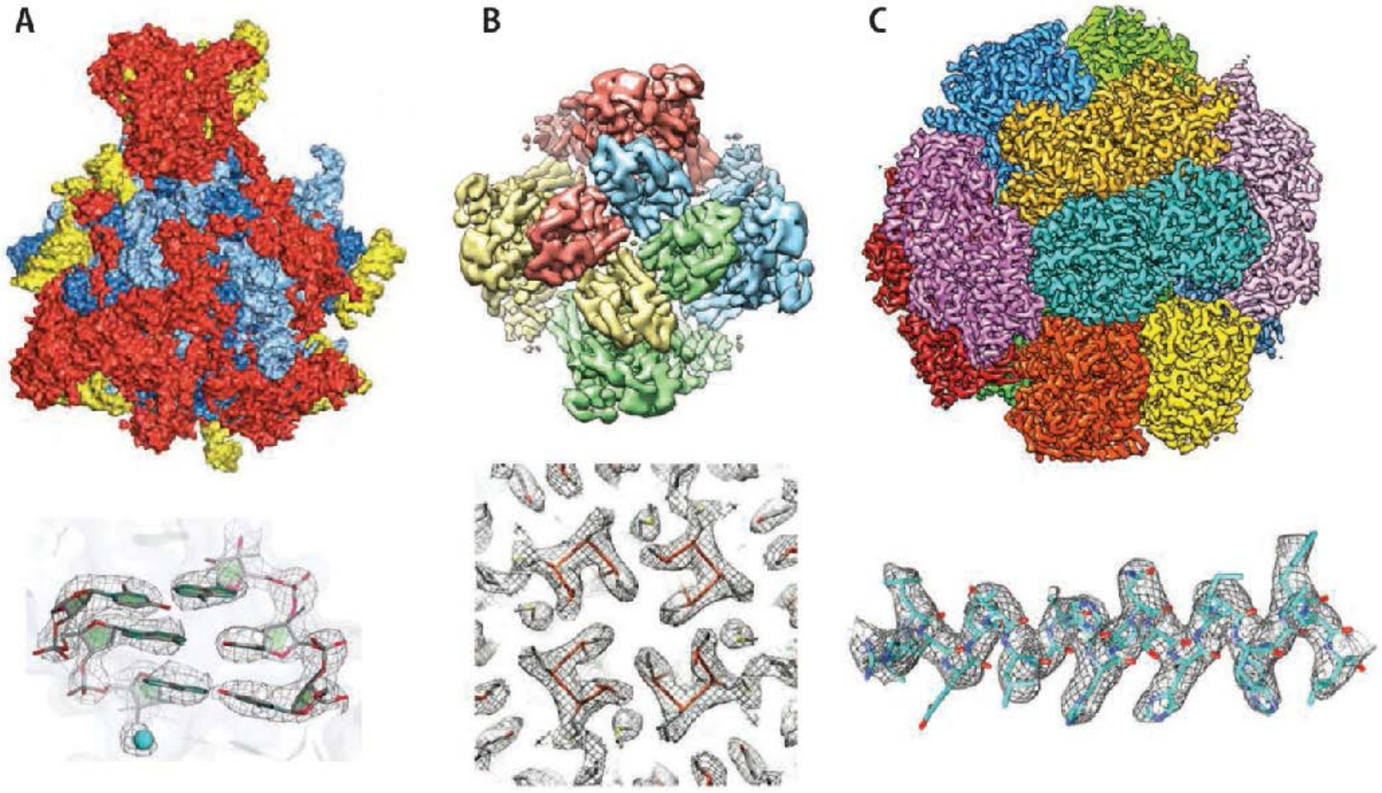


Sjors Scheres

Resolution revolution

- Dose fractionation image acquisition and motion correction become standard procedures.

- Direct detection camera is being used to produce a number of near atomic resolution reconstructions: “Resolution Revolution”



Yeast mitochondrial ribosome, 3.2Å

rat TRPV1 ion channel, 3.4Å

F420-reducing hydrogenase, 3.4Å

Werner Kuhlbrandt
“The Resolution Revolution”,
Science (2014)

Technologies that facilitated resolution revolution

- Direct electron detection camera (since 2012):
  - Single electron counting significantly improves detective quantum efficiency (DQE);
  - High frame rate enables dose fractionation and correction of beam induced image motion;
- New image processing algorithm based on maximum likelihood approach (first introduced by Fred Sigworth):
  - Facilitates better classification of good and “bad” particles;
  - Facilitates higher resolution structure determination;
- Modernization of electron microscope technologies:
  - Automated high-quality data acquisition;
  - Pipelined image processing enabled on-the-fly image processing;